

Analisi citazionale e indicatori bibliometrici nel modello Open Access

di Antonella De Robbio¹
(3 giugno 2007, aggiornato il 24 luglio 2007,
aggiunta citazione in data 23 gennaio 2008)

*Usage statistics are the key metric in determining the value of electronic products and services.
And they are an essential metric for both users and information providers.*
(NFAIS Forum, October 2004
http://www.nfaiss.org/events/event_details.cfm?id=26)

English abstract

This article is focused on the bibliometric methods used for research evaluation, within the well known field of the bibliometrics which uses mathematical and statistical techniques to analyze the publications distribution patterns, and to explore their impact into the scholarly communities. While the bibliometric methods are used more often in the LIS, bibliometrics offers a wide range of applications in a variety of fields of knowledge. Since the research evaluation requires a double approach: quantitative (bibliometric analysis) and qualitative (peer-review, panel...), the fundamental idea is that the measurement of the scientific knowledge uses both the methods.

Though the IF or impact factor is the most well known bibliometric index, is not the only one suggested by ISI (now Thompson) to refer. In the Open Access world are spreading a wide number of initiatives and projects regarding the "bibliometrics" field, in which are being testing alternative solutions to the traditional IF, for example the Eigenfactor, the Hirsch index and its variations like the g-index, a-index, h-b-index, tools that mark a meaningful change in this fascinating field.

While the indicators "author-produced" are the citations, those "reader-produced" are usage data collected through webserver and linkresolver logs. A quantitative bibliometric indicator of new generation is the Usage Factor UF, complementary but non substitutive of traditional or alternative impact factor indicator. It will be necessary to organize the log data collection to share methods to obtain significant analysis. The article goes over some examples of new tools/projects for citational analysis in the Web area, among these the Web Impact Factor (WIF), and for the statistical analysis of logs and data, both bringing back to the new science known as metric of the web or Webometric.

Within the OAI framework there are many research projects of the webometric type, involving the intellectual output archived into the repositories and their impact within the constitutive processes of the scholarly communication.

The theory assuming that an open access article is more likely to be read and therefore cited is matter of international debates and object of several studies, aimed to analyze the effects of the open access model and of the download on the citational impact. The article closes with a critical report on some of the most important works

¹ L'autrice ringrazia: Riccardo Ridi, per i suggerimenti e per l'indicazione di alcuni lavori che mi erano sfuggiti in quanto non disponibili ad accesso aperto; Elisabetta Marinoni, per gli utili commenti; Gianluca Drago per la segnalazione di alcune risorse preziose; Jean-Claude Guedon, faro illuminante verso la via d'oro dell'Open Access, per le stimolanti riflessioni a seguito delle chiacchierate serali all'OAI5 al CERN di Ginevra, lo scorso aprile 2007; Adriana Valente per avermi regalato il suo mirabile e mirato testo sulla "trasmissione d'élite o accesso alle conoscenze" che mi ha originato numerosi stimoli di riflessione; i revisori del Bollettino AIB per la pazienza e in particolare Giovanni Solimine, e ... tutti i colleghi che hanno letto e leggeranno questo lavoro.

investigating the validity of this theory, from the original study of Lawrence, on Nature 2001, to the most recent works, the three postulates of Kurtz, the analysis run in 2006 by the Cornell's Library staff on the MathSciNet database, the study of Moed which imposed a completely new methodological level with the introduction of the "variable citation window".

Useful tools for the citational bibliometric analysis, on the statistical usage measurement of the researches for the successful data monitoring, should be in the hands of the research producers, not in the hands of those commercial oligopolies that can influence the investment strategies and the national and international policies for the research.

Abstract in Italiano

L'articolo è focalizzato sui metodi bibliometrici usati per la valutazione della ricerca, entro l'ambito di studio noto come bibliometria, la quale utilizza tecniche matematiche e statistiche per analizzare i modelli di distribuzione delle pubblicazioni e per esplorare l'impatto entro le comunità scientifiche. Mentre i metodi bibliometrici sono più spesso usati nel campo LIS, la bibliometria offre ampie applicazioni in altre aree del sapere. Poiché la valutazione della ricerca richiede un duplice approccio: quantitativo (analisi bibliometrica) e qualitativo (peer-review, panel, ...) l'idea di fondo è che la misurazione della produzione del sapere scientifico usi entrambi gli approcci in modo combinato. Sebbene l'Impact Factor o IF, sia l'indicatore bibliometrico più conosciuto, non è il solo indicatore proposto dall'ISI (ora Thompson) a cui fare riferimento. Nel mondo Open Access inoltre stanno nascendo una serie di iniziative e progetti che interessano il campo della "bibliometria" dove si stanno sperimentando soluzioni alternative al tradizionale IF, tra cui l'Eigenfactor, l'Indice di Hirsch e le sue varianti come il g-index, a-index, h-b-index, strumenti che segnano una svolta epocale in questo affascinante campo di indagine.

Se gli indicatori generati dall'autore sono le citazioni, quelli generati dai lettori sono i dati sull'utilizzo che possono essere raccolti tramite server web e linkresolver logs. Indicatore bibliometrico quantitativo di nuova generazione è l'Usage Factor UF, complementare e non sostitutivo a fattori di impatto tradizionali o alternativi all'IF. Sarà necessario normalizzare il lavoro di raccolta dei dati di log al fine della condivisione di metodi per ottenere analisi significative. L'articolo passa in rassegna alcuni esempi di nuovi strumenti e/ progetti per l'analisi citazionale nell'ambiente Web, tra cui il Web Impact Factor (WIF) e per l'analisi statistica dei log e dei dati, che si riconducono alla nuova scienza nota come metrica del web o Webometrica.

Entro il framework OAI numerosi sono i progetti di ricerca di tipo webometrico che coinvolgono le produzioni intellettuali depositate negli archivi e il loro impatto entro i processi che costituiscono la catena della comunicazione scientifica.

La teoria che ipotizza che un articolo ad accesso aperto ha maggiori possibilità di essere letto e quindi citato è argomento di accesi dibattiti internazionali e oggetto di numerosi studi che hanno lo scopo di analizzare l'effetto dell'Open Access e dello scarico sull'impatto delle citazioni. L'articolo si chiude con una rassegna critica di alcuni dei lavori più importanti che indagano sulla validità di questa teoria, a partire dallo studio originale di Lawrence apparso su Nature nel 2001 e fino ai più recenti lavori, passando dai tre postulati di Kurtz, all'analisi condotta nel 2006 sulla banca dati MathSciNet dai bibliotecari della Cornell e fino al rigoroso studio di Moed che ha imposto un livello metodologico del tutto nuovo con l'introduzione della "variable citation window".

Strumenti utili per analisi bibliometriche sulle citazioni, sulla misurazione statistica dell'uso delle ricerche da parte dell'utente/autore, per il conseguente monitoraggio dei dati, dovrebbero essere nelle mani di chi la ricerca la produce e non di pochi oligopoli commerciali che possono influenzare le strategie di investimento e le politiche scientifiche nazionali e internazionali.

Contenuti

- 1. La bibliometria**
- 2. Metodi di valutazione della ricerca**
- 3. Metodi bibliometrici**
- 4. Indicatori bibliometrici**
- 5. L'Impact Factor (IF)**
- 6. Principali difetti dell'Impact Factor**
- 7. Indici bibliometrici alternativi all'Impact Factor**
- 8. L'Eigenfactor (EI)**
- 9. L'Indice di Hirsch o H-Index e sue varianti**
- 10. La metrica del Web e il Web Impact Factor (WIF)**
- 11. Iniziative per la standardizzazione delle statistiche e per la misurazione dei dati di utilizzo**
- 12. L'Usage Factor (UF)**
- 13. Le nuove frontiere delle analisi bibliometriche**
- 14. Impatto citazionale e usi di fonti informative multiple**
- 15. Open Access e impatto citazionale: da Lawrence a Moed**
- 16. Craig e i tre postulati di Kurtz**
- 17. Osservazioni conclusive**
- 18. Riferimenti bibliografici**

1. La bibliometria

La **bibliometria** è un ambito di studio applicato alle varie discipline scientifiche, che comprende anche l'analisi bibliometrica, sviluppatosi negli ultimi vent'anni grazie alla disponibilità online di banche dati di grandi dimensioni. E' collocata entro un contesto più ampio di **scientometria**, la scienza per la misura e l'analisi della scienza. Storia della scienza e tecnologia, filosofia della scienza e sociologia della conoscenza scientifica sono campi correlati alla scientometria. In pratica, la scientometria è spesso fatta usando la bibliometria che si riferisce alla misurazione delle pubblicazioni scientifiche².

La scientometria moderna è principalmente basata sulle opere di Derek J. de Solla Price, considerato il padre della scientometria e di Eugene Garfield, fondatore delle teorie dell'Institute for Scientific Information ISI.

L'*International Society for Scientometrics and Infometrics* (ISSI)³ fondata nel 1993 è l'associazione per eccellenza dei professionisti nel campo. Si possono considerare pionieri della scientometria alcuni nomi importanti, membri dell'ISSI, tra cui Tibor Braun, Bertram C. Brookes, Leo Egghe (ideatore del G-Index), Eugene Garfield (ideatore dell'Impact Factor), Wolfgang Glänzel, Belver C. Griffith, Ben Martin, Robert K. Merton (scomparso nel 2003), Henk F. Moed, Michael J. Moravcsik, Vasily V. Nalimov, Francis Narin, Ronald Rousseau, Henry Small.

La scientometria è un campo di studio, molto specialistico e piuttosto complesso, che trascende dal semplice conteggio delle citazioni e che coinvolge varie figure e comunità in tutto il mondo, specialisti e ricercatori in bibliometria, società scientifiche, periodici, e i partecipanti ai forum e liste di discussione entro le varie discipline.

Il concetto di bibliometria⁴ è stato usato negli ultimi decenni in riferimento a una disciplina che utilizza tecniche matematiche e statistiche per analizzare i modelli di distribuzione dell'informazione, ed in particolare delle pubblicazioni. Più precisamente è un insieme di metodi usati per studiare o misurare testi e informazioni. Se il suo sviluppo ha avuto come presupposto l'elaborazione automatica delle bibliografie secondarie, confluite nelle banche dati, il suo notevole successo è dovuto alla ricerca, da parte della comunità scientifica mondiale, di parametri oggettivi per la misurazione, detti indicatori bibliometrici, e la conseguente valutazione, della scienza e degli scienziati.

La maggior parte dei concetti e della terminologia correntemente usata per definire gli indicatori bibliometrici è dovuta a Eugene Garfield che, negli anni settanta, mise a punto un nuovo strumento nel campo delle bibliografie secondarie scientifiche, il *Science Citation Index*, prodotto dall'Institute for scientific information (ISI) di Philadelphia, sebbene fosse stato ispirato, ancora negli anni cinquanta, dalle idee di Vannevar Bush⁵. L'idea originale consisteva nell'arricchire la descrizione di ogni articolo con i riferimenti bibliografici in esso contenuti, aggiungendo un campo etichettato come *cited reference* (riferimento citato) e uno come *times cited* (numero di citazioni ricevute) campi utili a valutare

²Uno dei periodici più importanti nel campo della scientometria è *Scientometrics*, ma ve ne sono altri di prestigiosi come il *Journal of the American Society for Information Science and Technology* (JASIST) il *Social Studies of Science*, o l'elettronico *Cybermetrics*. *International Journal of Scientometrics, Informetrics and Bibliometrics*

³ <http://www.issi-society.info/>

⁴ Per un approfondito studio storico sulle origini degli indici di citazione nel circuito della comunicazione scientifica si rimanda all'ottimo lavoro di Adriana Valente cap. 4. Gli indici di citazione nel circuito di organizzazione, selezione e comunicazione di conoscenza scientifica, pp.75-97 In *Trasmissione d'élite o accesso alle conoscenze? Percorsi e contesti della documentazione e comunicazione scientifica*. A cura di Adriana Valente; testi di Sveva Avveduto et al. Milano, Franco Angeli, 2002

⁵ Vannevar Bush, "As we may think", The Atlantic Monthly, luglio 1945, vol. 176, No. 1, pp. 101-108. Il testo in lingua originale: <http://www.ps.uni-sb.de/~duchier/pub/vbush/vbush.shtml>. Una traduzione italiana: <http://www.uniopen.it/FC/ssd/ipertesto/think.html>.

l'impatto, l'importanza o la qualità di un'opera di uno scienziato in generale, dimostrando quanto spesso e dove sia stato citato.

Le due misure bibliometriche più conosciute sono il numero di citazioni e il fattore di impatto o impact factor IF, il numero di citazioni ricevute nell'anno corrente agli articoli pubblicati nei due anni precedenti diviso per il totale del numero di articoli pubblicati negli stessi due anni. All'inizio il fattore di impatto era disponibile soltanto in microformato, in seguito, verso gli anni ottanta l'ISI lo migrò in formato CD-Rom. Nel 1997 compare la versione su Web. Ad oggi 5900 periodici di ambito scientifico e tecnologico e 1700 nel campo delle scienze sociali sono presenti nel *Journal Citation Report JCR* strumento che fornisce il fattore di impatto come indice bibliometrico per questi periodici.

L'idea di fondo ad oggi è che si possa e che si debba misurare la produzione del sapere scientifico attraverso il conteggio delle pubblicazioni su delle riviste di alta qualità, usando degli indicatori bibliometrici o adottando altre tecniche. Fino a qualche anno fa la pubblicazione di un lavoro su una rivista era il mezzo principale di comunicazione del sapere entro le comunità scientifiche e contemporaneamente era anche considerato il criterio stesso di valutazione del valore di una ricerca. Pubblicare un articolo equivaleva, e equivale ancora oggi, in un certo senso, ad una certificazione di qualità, soprattutto se la rivista ha un certo prestigio entro la comunità. Per questa ragione, alle volte, gli archivi aperti o Open Archive sono erroneamente visti e percepiti, come contenitori di materiale di dubbia qualità o come ambienti non certificati.

Va detto però che entro il framework OAI numerosi sono i progetti di ricerca per l'analisi bibliometrica che coinvolgono le produzioni intellettuali depositate negli archivi e il loro impatto entro i processi che costituiscono la catena della comunicazione scientifica. L'impatto è definito generalmente come ciascun cambiamento o risultato di un'attività.

Susan Wainwright⁶ definisce l'impatto come *“ogni cambiamento risultante da un'attività, un progetto, un'organizzazione. Sono inclusi effetti previsti ed imprevisi, positivi e negativi, a lungo termine e a breve termine”*. La valutazione della ricerca in definitiva è una delle funzioni fondamentali e pervasive nel processo di crescita della conoscenza scientifica e può essere spiegata come quel processo decisionale sul merito di una ricerca scientifica basato su criteri e norme.

2. Metodi di valutazione della ricerca⁷

Il rapporto tra scienza e tecnologia è un rapporto storicamente complesso ed articolato, che da sempre ha caratterizzato i meccanismi di diffusione della conoscenza, ci dice Adriana Valente⁸, sottolineando che *“l'evoluzione storica della documentazione e comunicazione scientifica esplicita alcuni tratti fondamentali dell'interazione tra scienza e società”*. Il particolare binomio scienza e tecnologia entro un quadro “sociale” è oggetto di immemori riflessioni da parte degli studiosi, a partire a Platone.

⁶ Susan Wainwright, *Measuring Impact – A guide to Resources*. NCVO, 2002

⁷ Per un lavoro completo sulla valutazione delle pubblicazioni a stampa fino all'era degli archivi aperti si veda il lavoro di Comba, Valentina (2003) *La valutazione delle pubblicazioni: dalla letteratura a stampa agli Open Archives*. *Bollettino AIB* 2003(1):pp. 65-76, valido contributo allo studio degli strumenti valutativi nel mondo della ricerca e della comunicazione scientifica in senso lato. Nell'articolo vengono attentamente considerate gli usi e le critiche all'Impact Factor dell'ISI oltre ad interessanti considerazioni in merito al ruolo del bibliotecario entro la bibliometria.
<http://eprints.rclis.org/archive/00000095/>

⁸ *Trasmissione d'élite o accesso alle conoscenze? Percorsi e contesti della documentazione e comunicazione scientifica*. A cura di Adriana Valente; testi di Sveva Avveduto et al. Milano, Franco Angeli, 2002.

La valutazione della ricerca richiede un duplice approccio: **quantitativo** (ossia in termini numerici di impatto scientifico, brevettabilità, della presenza di contratti con aziende interessate a temi di ricerca); **qualitativo**, il giudizio dei pari (valutatori) o peer-review, che risulta a tutt'oggi il più importante dei metodi per una valutazione significativa della qualità. Appare evidente come sarebbe necessario associare ai metodi quantitativi uno o più metodi qualitativi.

Oltre all'analisi bibliometrica (quantitativa), usata un po' ovunque, in altri paesi invece vengono usati più metodi, tra i quali il *panel* (modello anglosassone), il *peer review* (il giudizio dei pari) e l'analisi del beneficio economico.

Nel mondo universitario italiano, è utilizzata anche una classificazione gerarchica - rideterminata sulla base del Decreto Ministeriale pubblicato su G.U. n. 249 24 ottobre 2000 (supplemento ordinario 175)⁹ - basata su quattordici macro-settori disciplinari, identificati dalle lettere A-I, K-N, P, Q, S e V. suddivisi a loro volta in 441 settori scientifico-disciplinari. In tale sistema la valutazione della produzione scientifica viene proposta la disaggregazione e la ponderazione degli scritti, in relazione ai suddetti settori disciplinari.

L'**analisi bibliometrica** è una branca specifica della bibliometria e si basa sulla premessa che il lavoro di un ricercatore ha valore solo quando è giudicato dai pari aventi merito scientifico. E' focalizzata nell'analisi quantitativa di un'unità della comunicazione scientifica, intesa come un articolo, capitolo di volume, o altro tipo di paper di ricerca. Se una ricerca pubblicata in una rivista qualificata è stata approvata dai *referee* o da una *peer review* (da un comitato dei pari) e quindi dall'editore, il lavoro pubblicato verrà citato da altri ricercatori in riconoscimento al contributo che esso ha dato al settore. Recentemente le analisi bibliometriche si stanno conducendo anche entro gli archivi aperti con risultati assai sorprendenti. Ne conseguirebbe che più volte un lavoro viene citato, maggiore è la sua qualità.

Il merito principale dell'analisi bibliometrica, ma anche il suo limite, è la sua natura quantitativa, sebbene essa si correli con il metodo qualitativo della *peer review*, fornendo nell'insieme risultati discretamente coerenti.

Le critiche all'analisi bibliometrica sono prevalentemente critiche al fattore di impatto, come vedremo più avanti.

Il **panel** richiede che la valutazione sia basata su visite da parte di gruppi valutatori (*referee panels*) delle istituzioni da valutare ed incontri con il personale impegnato nella ricerca, con gli studenti ed il personale tecnico. Le tipologie di presenza dei panels sono estremamente diverse per ogni Ateneo e rispondono ad esigenze locali. Si tratta di un metodo di valutazione usato dalla britannica RAE Research Assessment Exercise¹⁰.

La **peer review**, tipica del mondo accademico, implica l'impiego, da parte dell'organismo che esegue la valutazione, di valutatori esterni all'istituzione da valutare e, il giudizio dei pari, risulta tutt'oggi il più importante dei dati che permettono una valutazione significativa della qualità. La *peer review* nacque nel 1665 con le *Philosophical transaction* della Royal Society of London e fu scelta come meccanismo per l'allocatione delle risorse per il finanziamento alla ricerca nei Stati Uniti negli anni tra il quaranta e il cinquanta.

Recentemente il modello della peer-review è pesantemente posto in discussione da parte di alcune comunità. L'interessante dibattito su *Nature* propone 22 contributi che analizzano il sistema e offrono

⁹ http://www.miur.it/0002Univer/0021Offert/0092Settor/index_cf2.htm

¹⁰ <http://www.rae.ac.uk/panels/>

una panoramica sul tema, presentando diverse posizioni e nuove prospettive e affrontando tutti i diversi aspetti, etici, tecnologici e pratici¹¹.

Le diverse sezioni del dibattito pongo in luce gli aspetti più caldi:

- ❑ Certification in a Digital Era¹²
- ❑ Evolving Peer Review for the Internet¹³
- ❑ An Open, Two-Stage Peer-Review Journal¹⁴
- ❑ Opening Up the Process¹⁵
- ❑ Reviving a Culture of Scientific Debate¹⁶
- ❑ Wisdom of the Crowds¹⁷

Interessante è il nuovo modello proposto da Nature Precedings¹⁸, iniziativa del gruppo Nature, quale luogo aperto per i ricercatori di condivisione dei lavori di ricerca pre-print, non pubblicati, manoscritti, presentazioni, posters, white papers, lavori tecnici, scoperte supplementari e altri documenti del circuito della comunicazione scientifica. La sottomissione a Nature Precedings di tali tipologie di documenti è soggetta a controllo formale ma non soggetta a peer review. I campi interessati sono la biologia, la medicina (esclusi i trial clinici); la chimica e le scienze della terra sul modello del server ArXive dei fisici. In sostanza viene “spostato” in avanti, il momento del referaggio, che avviene su larga scala, da parte di tutta la comunità compresi gli editori, di modo che non vi sia a priori una selezione di “pochi” (comitato dei pari) come avviene nel modello tradizionale di peer review. Timo Hannay, direttore del Web Publishing di Nature, in una lettera apparsa di recente su Tim O'Reilly Radar¹⁹, spiega le ragioni della creazione di Nature Precedings *“The traditional way for scientists to share their research results is through journals. These have the benefit of being peer-reviewed, citable and archival, but as a communication channel they are also relatively slow and expensive. [...] Because Nature Precedings isn't peer-reviewed (to be more accurate, the submissions are subjected to open review *after* their release, through user comments and votes), we see it as complementing rather than competing with traditional journals, just as arXiv.org operates alongside the peer-reviewed journals in physics”*

In Nature Precedings i contenuti sono rilasciati sotto licenza Creative Commons e ciascun item è reso citabile usando codici DOI o Handle, stessi sistemi usati per i lavori scientifici sottoposti al tradizionale processo di peer-review..

L'analisi del beneficio economico non è altro che un'**analisi costi-benefici** al fine di valutare la convenienza finanziaria ed economica di progetti di ricerca, non solo dove siano intervenuti investimenti da parte di aziende private, ma soprattutto per quei progetti finanziati con fondi pubblici. Nella valutazione finanziaria si calcola la convenienza del progetto in termini di investimento considerando le variabili finanziarie come i costi e i ricavi monetari. La valutazione economica stima il valore monetario in termini di efficacia o di impatto esterno del progetto come investimento, considerando fattori sociali (negativi o positivi) come il miglioramento o degrado dell'ambiente, aumento dell'occupazione, aumento del traffico. Per questa ragione è più facile applicare analisi di

¹¹ http://blogs.nature.com/nature/peerreview/debate/comments/2006/06/technical_solutions_certificat.html

¹² <http://www.nature.com/nature/peerreview/debate/nature05008.html>

¹³ <http://www.nature.com/nature/peerreview/debate/nature04997.html>

¹⁴ <http://www.nature.com/nature/peerreview/debate/nature04988.html>

¹⁵ <http://www.nature.com/nature/peerreview/debate/nature04994.html>

¹⁶ <http://www.nature.com/nature/peerreview/debate/nature05005.html>

¹⁷ <http://www.nature.com/nature/peerreview/debate/nature04992.html>

¹⁸ <http://precedings.nature.com/>

¹⁹ http://radar.oreilly.com/archives/2007/06/nature_precedin.html

beneficio economico per progetti di ricerca che hanno ricadute di sviluppo economico e di conseguenza sociale, nel campo dell'industria (Ricerca e Sviluppo), in particolare bioscienze, medicina, ingegneria, meccanica, statistica...

Un altro criterio importante che si può aggiungere ai precedentemente segnalati, è quello della percentuale di **internazionalizzazione delle pubblicazioni**: questo parametro però, risulta rilevante solo per alcuni settori di attività scientifica.

L'indice o **grado di internazionalizzazione** è la percentuale di lavori che vengono pubblicati su giornali internazionali, da singoli ricercatori, da gruppi di ricerca o da enti o istituzioni e comprende anche lavori che rientrano sotto brevetti o i contratti con aziende internazionali.

E' un dato di fatto che un buon 90% dei lavori che sono stati pubblicati in riviste scientifiche accademiche non sono mai stati citati. In realtà va detto anche che quasi un 50% dei lavori pubblicati non sono mai stati letti da nessuno oltre che dai propri autori, i referi e i curatori del periodico nel quale l'articolo è stato pubblicato²⁰.

3. Metodi bibliometrici

La bibliometria offre ampie applicazioni in numerose aree del sapere dove si usano metodi bibliometrici per esplorare l'impatto entro le comunità scientifiche di un insieme di ricercatori o di un particolare lavoro scientifico. Storicamente i metodi bibliometrici sono stati usati per tracciare le relazioni tra le citazioni dei periodici accademici, per questa ragione i metodi bibliometrici - più che la bibliometria in senso lato - rientra nei campi della *LIS Library and Information Science* e discipline correlate.

L'analisi citazionale e **l'analisi di contenuto** sono metodi bibliometrici comunemente usati dalla bibliometria.

Altre applicazioni bibliometriche includono: la creazione di tesauri, la misurazione della frequenza di termini, l'esplorazione delle strutture grammaticali e sintattiche dei testi.

L'analisi citazionale è lo strumento cardine della disciplina di ricerca nota come bibliometria.

L'analisi citazionale è il metodo più noto e usa citazioni nelle produzioni intellettuali scientifiche per stabilire connessioni ad altri lavori o altri ricercatori. Accoppiamenti di co-citazioni e accoppiamenti bibliografici sono modelli specifici dell'analisi citazionale.

L'analisi citazionale è l'esame della frequenza e del modello di citazioni negli articoli e/o testi in generale. A causa di una crescita senza precedenti della disponibilità di risorse elettroniche (e-resources) una delle questioni emerse con forza e oggetto di indagine da parte degli studiosi è quanto spesso le risorse elettroniche sono citate nei singoli campi di interesse. Per esempio ci sono studi che registrano un alto tasso di citazione per lavori online nel campo della letteratura informatica, ciononostante per l'ambito umanistico, la versione a stampa è ancora un punto di riferimento.

Michael Koenig della *Palmer School of Library and Information Science* all'Università di Long Island di New York è uno tra i sostenitori dell'analisi citazionale in quanto afferma che le citazioni forniscono – a dispetto delle loro ambiguità – “una misura oggettiva di ciò che è variamente chiamato

²⁰ Renato Spigler, "Peer-Reviewing and Electronic Publishing", High Energy Physics Libraries Webzine, issue 6, March 2002 URL: <<http://library.cern.ch/HEPLW/6/papers/5/>>

produttività, significanza, qualità, utilità, influenze, efficacia, o impatto degli scienziati e delle loro produzioni scientifiche”²¹.

L’analisi citazionale è considerata la macchina dove poggia il fattore d’impatto, noto come IF. Sebbene per molti decenni il *Science Citation Index* dell’ISI Institute for Scientific Information - ora *Web of Science WoS* di Thompson ²² è stato considerato lo strumento principe per la misurazione delle citazioni, da qualche tempo i servizi Web stanno mettendo in discussione il predominio non solo degli strumenti dell’ISI ma dello stesso IF.

Poiché da studi effettuati negli ultimi anni è emerso che ciascun servizio produce risultati anche piuttosto differenti l’uno dell’altro, in termini di analisi quantitativa appare quanto mai necessario usare diverse fonti di citazione per giudicare il reale impatto di un’opera scientifica. Il Web infatti ha avuto un enorme impatto sulla ricerca dell’analisi citazionale. Negli anni recenti dozzine di database tra cui *Scopus*²³ e anche *Google Scholar*²⁴ sono comparsi sulla scena, permettendo modalità e modelli di citazione dei lavori scientifici mai visti in precedenza²⁵. Questo potrebbe segnare, secondo alcuni, la fine del monopolio sull’analisi citazionale durato oltre quarant’anni e detenuto dal commerciale statunitense Thomson Scientific, meglio conosciuto con il nome precedente di ISI. Gli scienziati hanno spesso criticato i database citazionali di ISI per il limitato numero di titoli di periodici indicizzati i quali erano quasi esclusivamente di lingua inglese e provenienti principalmente dal Nord America e dall’Europa Occidentale, inoltre non erano coperti né volumi monografici né gli atti delle conferenze. Il Web sta conducendo a numerosi nuovi metodi di misurazione delle citazioni, precedentemente impraticabili, come il conteggio degli articoli scaricati, l’analisi dei link, il *Google’s Page Rank*, il *Web citation* e gli indici bibliometrici alternativi come l’H-index o simili.

Se da una parte il migrare dei database dell’ISI su Web tramite lo strumento *WoS* aumentò notevolmente la potenza nel campo dell’analisi citazionale e il valore degli stessi strumenti ISI, d’altro canto lo stesso Web cominciava a produrre nuovi strumenti e nuovi luoghi di pubblicazioni elettroniche in aperta competizione e sfida al modello ISI che dimostrava ormai una certa obsolescenza, soprattutto continuando ad ignorare le pubblicazioni anche prestigiose del mondo Open Access, i lavori depositati negli archivi aperti e i contenuti delle pagine web personali degli autori accademici, tutto materiale spesso di alta qualità messo a disposizione immediatamente dopo la sua creazione per un’ampia disseminazione entro le comunità a livello globale.

²¹ Koenig, M.E.D. (1983). Bibliometric indicators versus expert opinion in assessing research performance. *Journal of the American Society for Information Science*, 34, 136–145.

²² Il *Science Citation Index (SCI)* fu lanciato nel 1964 e ben presto divenne popolare tra scienziati e bibliotecari e tutt’oggi è uno dei database multidisciplinari più importanti del mondo. SCI è affiancato da altri database citazionali come l’*Arts and Humanities Citation Index (A&HCI)* e il *Social Science Citation Index (SSC)*. Insieme ad oggi contano circa 40 milioni di record da oltre 8700 periodici scientifici tra i più prestigiosi del mondo. Fino al 1988 questi indici esistevano solo in forma cartacea a stampa, sebbene già dalla metà degli anni settanta fossero possibili ricerche online usando sistemi di information-retrieval messi a disposizione da soggetti terzi come Dialog. Dal 1988 comparirono le prime edizioni su CD-Rom e finalmente nel 1997 i database ISI migrarono su web con il lancio del *Web of Science (WoS)*.

²³ Base dati bibliografica multidisciplinare prodotta da Elsevier e lanciata nel 2004. <http://www.info.scopus.com>. Ad oggi comprende registrazioni bibliografiche ricavate dagli spogli di oltre 15.000 periodici peer-reviewed (di cui 531 open access), pubblicati da 4.000 editori scientifici, tecnici e medici: copre circa l’80% della produzione scientifica mondiale ed è la più ricca al mondo con un totale di più di 27 milioni di record; la copertura retrospettiva è relativa agli ultimi 40 anni. E’ incluso Citation Tracker, nuovo strumento per la ricerca delle citazioni bibliografiche e l’H-index come nuovo indicatore bibliometrico.

²⁴ <http://scholar.google.com/schhp?hl=it>

²⁵ Bakkalbasi, Nisa, Kathleen Bauer, and Janis Glover, et al (2006), Three Options for Citation Tracking: Google Scholar, Scopus and Web of Science
<http://eprints.rclis.org/archive/00006080/>

L'analisi di contenuto invece, nota anche come **analisi testuale** quando condotta esclusivamente su testi, è una metodologia standard nel campo delle scienze sociali applicata per lo studio del contenuto della comunicazione. Fu definita da Earl Babbie come “*lo studio di comunicazioni umane registrate*”, laddove Babbie per comunicazioni umane intende libri, siti web, disegni e leggi.

Harold Lasswell formulò la questione centrale dell'analisi di contenuto con la frase “*Chi dice cosa, a chi, perché, in quale contesto e in che modo e con quale effetto*”.

Il metodo dell'analisi di contenuto mette in condizione i ricercatori di includere grandi quantità di informazioni testuali e di identificare sistematicamente le loro proprietà, per esempio le frequenze delle parole chiave più usate (*KWIC KeyWord In Context*) per individuare le strutture più importanti del loro contenuto comunicativo. Tali aggregazioni di informazioni testuale dovrebbero poi essere categorizzate in accordo a determinati framework teorici di riferimento, i quali informano i sistemi di data analysis fornendo alla fine una lettura con significato logico del contenuto analizzato.

In tale contesto per analisi dei dati o *data analysis* si intende l'atto di trasformare un dato con lo scopo di estrarne informazioni utili e facilitare conclusioni che abbiano un significato logico. In relazione al tipo di dato e alla domanda, questo potrebbe includere applicazioni di metodi statistici, *curve fitting*²⁶, selezione e scarto di determinati sotto-insiemi basati su criteri specifici, o altre tecniche. Rispetto al *data mining*, l'analisi dei dati è solitamente intesa non tanto come atto focalizzato alla scoperta di modelli impreveduti nascosti nei dati, ma alla verifica ai fini di un'approvazione o rifiuto di un modello attuale, o all'estrazione dei parametri necessari per adattare un modello teorico sperimentale alla realtà.

4. Indicatori bibliometrici

Accanto all'uso di strumenti di ricerca propri di ogni disciplina, gli indicatori bibliometrici sono utilizzati in misura crescente nelle politiche dei comitati o delle agenzie di valutazione della ricerca per valutare la qualità e la produttività di istituzioni di ricerca e quindi per orientare i finanziamenti pubblici.

A causa delle assunzioni su cui si basano (ad esempio citazione = impatto) e della limitazione dei dati disponibili nelle indagini su banche dati tradizionali, dovrebbero essere utilizzati con cautela.

La formula dell'Impact Factor, fin dal suo sorgere, consentì una grande quantità di elaborazioni statistiche, che hanno portato la bibliometria ad assumere il ruolo di disciplina sempre più sofisticata ed applicabile ai più svariati contesti. Mediante l'*analisi citazionale*, si può quindi valutare il *fattore d'impatto* IF di un articolo, di un autore, di una rivista, di un gruppo di ricercatori, di una nazione, tenendo anche conto di vari *indicatori di obsolescenza* (sincronica e diacronica).

Solitamente si confonde spesso tra i casi in cui si fa uso di dati specifici delle pubblicazioni esaminate desunti dal *Science Citation Index* o in casi in cui si fa riferimento all'IF medio della rivista dove l'articolo è pubblicato, ricavato dal *Journal of Citation Index* dell'ISI.

Gli indicatori utilizzati nell'analisi bibliometrica proposta dall'ISI non si limitano al solo Fattore di Impatto, ma si differenziano in vari indicatori bibliometrici²⁷:

²⁶ trovare una curva sovrapponibile ad una serie di punti dati e altri possibili vincoli

²⁷ <http://scientific.thomson.com/media/scpdf/ws7-1005-q-it.pdf>

- ❑ **Immediacy Index:** misura quanto successo sta avendo il lavoro nell'anno di pubblicazione e in relazione a quanto velocemente un articolo della rivista è mediamente citato e quanto spesso gli articoli della rivista sono citati nello stesso anno;
- ❑ **Cited Half Life:** ciclo di emivita di un lavoro scientifico; misura la validità nel tempo degli articoli citati o la durata delle citazioni nel tempo. Misura il numero degli anni, andando all'indietro da quello corrente, in cui si raggiunge il 50% delle citazioni totali ricevute dalla rivista nell'anno presente.
- ❑ **Rate of Cites Index:** rappresenta un indice di qualità del singolo lavoro, basato sull'assioma che quanto più il lavoro è citato da altri ricercatori tanto più rilevante è il suo valore scientifico.
- ❑ **Citation Impact:** è calcolato per uno specifico soggetto o autore o istituzione o paese sulla base del rapporto tra il numero di citazioni ricevute e il numero di articoli pubblicati (un elevato numero di citazioni indica che la pubblicazione ha avuto un forte impatto). riflette la misura in cui un determinato articolo è stato utilizzato per successivi lavori scientifici e quindi, in qualche misura, il suo impatto sullo sviluppo della scienza. Questi dati si prestano a diverse analisi:
 - studi di produttività scientifica attraverso il conteggio delle pubblicazioni per autore o per istituzione, in rapporto alle risorse (pubblicazioni per ricercatore);
 - studi sulle collaborazioni scientifiche: ad esempio il numero di articoli firmati da autori di più paesi oppure da ricercatori di università e di industrie; l'esame delle citazioni permette anche di mettere in evidenza gruppi di ricerca che collaborano e che si citano mutualmente in modo molto frequente;
 - analisi della struttura di un campo scientifico con l'identificazione di parole-chiave che si ripetono o che vengono usate congiuntamente; è così possibile individuare nuovi fronti di ricerca caratterizzati da un approccio e da una terminologia comune.

Tale misura presenta alcuni limiti:

- presenza e numerosità delle citazioni;
- distribuzione temporale delle citazioni;
- concentrazione delle citazioni in pochi lavori o dispersione in molti;
- natura delle citazioni (positive/negative);
- qualità e numero complessivo delle riviste dalle quali sono desunte le citazioni;
- predominanza di riviste scritte in lingua inglese.

5. L'Impact Factor²⁸ (IF)

L'impact Factor (IF): misura il livello della ricerca scientifica, su scala nazionale ed internazionale, delle pubblicazioni scientifiche. Varia moltissimo da area ad area; è, infatti, una misura del numero di citazioni dei lavori pubblicati in una certa rivista rispetto al numero totale di lavori pubblicati dalla stessa rivista negli anni precedenti; risulta, quindi, pesato anche per la diffusione del giornale in questione; è ovvio che, essendo maggiore la diffusione in certi campi, c'è una certa disparità fra le diverse discipline, superata dal fatto che non si effettuano confronti tra settori ma all'interno di un dato settore.

²⁸ Eugene Garfield, The History and Meaning of the Journal Impact Factor. JAMA, January 4, 2006—Vol 295, No. 1 (Reprinted). <http://www.garfield.library.upenn.edu/papers/jamajif2006.pdf>

Calcolo dell'Impact Factor.

Se nel corso del 2006 gli articoli usciti nel 2005 e nel 2004 su una rivista **X** sono stati citati rispettivamente 260 e 95 volte, e se il totale degli articoli pubblicati su quella rivista nel 2005 e nel 2004 è 80, allora l'*IF* della rivista X è dato da:

$$IF(X) = \frac{260 + 95}{80} = 4.43$$

A livello internazionale l'Impact Factor è considerato l'indicatore del grado di penetrazione dei risultati della ricerca nella comunità scientifica e della loro considerazione da parte di altri ricercatori.

La letteratura sull'IF vede due fronti contrapposti: da una parte quelli che lo criticano duramente mettendo in evidenza i limiti ed i difetti di tale indicatore bibliometrico e, d'altra quelli che ne esaltano l'utilità, nonostante tutto. Fino a qualche tempo fa veniva sottolineato il fatto che gli elementi positivi prevalgono su quelli negativi, per il largo uso che se ne fa a livello nazionale ed internazionale, ma anche per la mancanza di indicatori migliori non basati sull'autoreferenzialità. Come vedremo in seguito negli ultimi due o tre anni stanno sorgendo indici alternativi. Va detto che purtroppo sono gli stessi ricercatori/scienziati i responsabili di queste lacune, proprio perché considerati gli strumenti e le conoscenze che alcune comunità hanno a disposizione - per background culturale e per interesse diretto nel trovare nuovi modi più adeguati - l'uso di indicatori poco conformi per lunghi decenni non fa altro che procrastinare tali distorsioni non di certo vantaggiose per il progresso scientifico e tecnologico.

In ogni caso a favore dell'IF depongono due condizioni:

- l'aspetto burocratico: chi deve assegnare fondi a seguito di richieste da parti di gruppi di ricerca non deve necessariamente conoscere la disciplina, né le riviste scientifiche di riferimento. Il semplice parametro numerico dell'IF permette il raffronto interno al settore che si considera.
- risulta comunque un metodo accettabile in comunità scientifiche che non sono in grado di autovalutarsi criticamente.

Ma quali sono i criteri a monte che determinano quando e come una rivista rientri nel prestigioso WoS e venga dotata di Impact Factor? Nell'editoriale di JoP Journal of the Pancreas del maggio 2007²⁹ i curatori della rivista raccontano delle difficoltà di ottenere l'IF da Thompson, nonostante la rivista sia dotata di comitato internazionale e di autori tra i più prestigiosi che scrivono sulla rivista, sia tra le più quotate del settore (pancreatologia), sia indicizzata dai grandi sistemi di I&A quali PubMed, EMBASE, Scopus e abbia un tasso di rifiuto degli articoli superiore al 45%. Nonostante questo la rivista non ha un IF e alle varie richieste alla Thompson non ne è mai stata data risposta o spiegato il motivo del rifiuto. Gli autori hanno analizzato il numero di citazioni secondo il modello ISI di Garfield, durante i sette anni di vita della rivista. E' stato quindi stimato un possibile fattore di impatto del periodico che è risultato pari a 0.846 per il 2005 e 0.886 per il 2006. L'indice di immediatezza per il 2005 è stato stimato pari a 0.398 e per il 2006 a 0.691. Non sono valori alti ma se si va ad analizzare in JCR quante

²⁹ Pezzilli R, Morselli-Labate AM, Morotti L, Steering Committee of "JOP. Journal of the Pancreas: Considerations on a 'Pancreatic' Electronic Journal at the Beginning of Its 8th Year of Publication"

JOP. J Pancreas (Online) 2007; 8(3):263-267.

Editorial published by the Steering Committee of JoP

<http://www.joplink.net/prev/200705/index.html>

riviste sono presenti con valori più bassi di quelli calcolati per JoP è sorprendente notare che per il solo fattore di impatto le riviste che hanno IF inferiore sono 41.7% e II ben il 78.7%. Forse l'esclusione è dovuta al fatto che si tratta di una rivista ad accesso aperto?

6. Principali difetti dell'Impact Factor

L'IF è un parametro di riferimento che consente di stilare graduatorie delle riviste per ciascuna delle categorie disciplinari adottate dal *JCR, Journal Citation Report*. Anche se generalmente accettata, l'assunzione che ad un'elevata frequenza di citazione corrisponda un'elevata qualità della rivista, sul piano concettuale, ha scarso fondamento. Poiché l'IF viene usato come parametro di riferimento per stilare graduatorie delle riviste più prestigiose, tale pratica ha generato una serie di critiche, in particolare da parte di alcuni matematici, che hanno portato a definire l'IF come uno strumento di promozione commerciale o meglio di auto-promozione commerciale da parte dell'ISI. Alessandro Figà Talamanca nel suo storico discorso al convegno SINM del 2000³⁰ scrive “... oltre alla decisione elementare di escludere le riviste che non sono sufficientemente regolari, l'ISI cercò di sviluppare criteri che le consentissero di misurare l'utilità, ai suoi fini, dell'inclusione o esclusione di una rivista. [...] Il risultato finale di questa evoluzione, cui l'ISI ha contribuito negli ultimi anni, è stato quello di favorire i grandi editori commerciali, a scapito delle pubblicazioni legate ad istituzioni scientifiche, specialmente quelle dei paesi europei non di lingua inglese, di far lievitare i costi delle pubblicazioni” e relativamente all'IF dice “è possibile presentarlo alle biblioteche scientifiche delle università e centri di ricerca americani, come uno strumento per decidere se vale la pena di acquisire o mantenere l'abbonamento ad una rivista. Queste biblioteche sono, come si è detto, i principali clienti dell'ISI ed era ed è opportuno convincerle che la banca dati dell'ISI è completa ed affidabile”

Le critiche mosse all'IF si possono suddividere in alcune aree di ben definite.

La prima si riferisce alle variazioni tra campi disciplinari cui la rivista afferisce (settori disciplinari generici presentano, in media, fattori di impatto più elevati rispetto agli specialistici). Alcune discipline, per esempio quelle cosiddette trasversali, sono caratterizzate da una produzione di letteratura relativamente inferiore ad altre. Infatti sempre Figà Talamanca osserva che il massimo IF per le riviste matematiche è 2,6 e che solo 10 riviste sulle 265 elencate hanno un IF di almeno 1,5. Tra le dieci prime riviste della lista ce ne sono tre di indubbia qualità, altre che sono entrate abusivamente in quanto si occupano di discipline affini alla matematica, altre tre sono di matematica applicata, riviste senza dubbio probabilmente molto citate fuori dell'ambiente matematico. “In genere abbondano tra le prime 20 riviste quelle parzialmente estranee alla matematica o quelle che si occupano di matematica applicata [...] o che si dedicano ad un argomento estremamente specialistico e molto alla moda.”

Altre discipline dispongono di pochi periodici recensiti a livello internazionale. Ne consegue una frequenza di citazione intrinsecamente bassa, ossia un basso valore dell'IF. Ma relativamente al Paese, Figà Talamanca prosegue “La matematica russa risulterebbe pressoché inesistente se lo IF dovesse essere preso come indice di qualità. Considerazioni analoghe, anche se non così forti e precise, possono essere fatte sulla matematica francese e su quella giapponese. Ne segue che, per la matematica, la considerazione dello IF non è affidabile nemmeno nella valutazione complessiva della

³⁰ Alessandro Figà-Talamanca, *L'Impact Factor nella valutazione della ricerca e nello sviluppo dell'editoria scientifica*. IV Seminario Sistema Informativo Nazionale Per La Matematica SINM 2000 : un modello di sistema informativo nazionale per aree disciplinar. Lecce, Lunedì 2 ottobre 2000
<http://siba2.unile.it/sinm/4sinm/interventi/fig-talam.htm>

ricerca di un intero paese [...] E' improbabile che questa inaffidabilità si riferisca solo alle discipline matematiche e non investa invece almeno le altre discipline scientifiche, fuori del settore biomedico".

La seconda area di miglioramento dell'IF si riferisce alla dimensione della rivista e la terza alla differenza di impatto tra tipi di riviste e di articoli in esse contenute.

La quarta area critica è relativa alla politica editoriale della rivista, la sua periodicità e la capillarità nella distribuzione hanno riflessi significativi sull'IF, di conseguenza si possono riscontrare notevoli oscillazioni dei valori di IF da un anno all'altro per una stessa rivista. L'irregolarità nella pubblicazione fa sì che riviste anche di ottimo livello scientifico non siano incluse nelle liste dell'ISI e quindi rimangano prive di IF. La lingua è un altro fattore: le riviste scientifiche in lingua inglese sono più diffuse (e, quindi, dotate di IF maggiore) di quelle editate in altre lingue; spesso, poi, quelle editate negli USA lo sono di più di quelle editate in Europa ed in Giappone. La limitazione dell'intervallo temporale considerato nella definizione dell'IF può risultare penalizzante per le riviste dotate di procedura di 'refereeing' particolarmente severa, rigorosa e lunga.

Infine l'asimmetria nelle componenti al numeratore (dove si includono le citazioni provenienti da qualsiasi tipo di pubblicazione) e al denominatore (dove per pubblicazione si considerano solo articoli, note, atti di convegni e recensioni) che individua una certa disomogeneità tra aree in relazione alle citazioni di tipologie di materiale diverso: in alcune aree disciplinari il numero delle citazioni di testi e monografie è più elevato che in altre: tale fattore riduce l'entità delle citazioni che entrano nel computo dell'IF. Esistono inoltre differenze, anche significative, nel numero medio di riferimenti bibliografici indicati nell'*articolo medio* delle varie aree.

In aggiunta alle osservazioni di Figà Talamanca e dei matematici, ve ne sono altre provenienti da altre comunità e che sono state raccolte dai vari studi di settore³¹. Dall'ambiente umanistico emerge che alcuni settori disciplinari hanno per oggetto di studio argomenti di interesse tipicamente locale: di conseguenza sono caratterizzate da valori di IF bassi. Il computo dell'IF non distingue tra citazioni bibliografiche ed autocitazioni, e tra citazioni 'positive' e citazioni 'negative'. Gli articoli di rassegna o recensione sono citati molto più frequentemente di quelli originali; pertanto, le *reviews* godono dei valori più elevati di IF, così come le riviste di carattere divulgativo, e quelle multidisciplinari sono citate più delle altre.

In conclusione, il criterio del numero delle citazioni in riviste scientifiche e dell'IF oltre ad essere praticati poco e solo per alcune aree disciplinari come la medicina e alcuni settori dell'ingegneria, devono essere usati con cautela perché rischiano di scambiare la quantità per la qualità. Nella relazione sullo stato delle Università Italiane 2006, il Presidente della CRUI Conferenza dei Rettori delle Università Italiane. Prof. Guido Trombetti, così scrive a proposito dell'IF quale strumento di valutazione della ricerca *"L'Impact Factor e il Citation Index possono andare bene per la Fisica. Sono assolutamente inapplicabili, per il momento almeno, in molte aree umanistiche"*³².

Non esiste, infatti, un metodo o un elenco di indicatori che sia valido in assoluto dato che non si può prescindere dalla qualità della ricerca e dalle sue finalità nel comparto economico-produttivo nazionale. Se si utilizzano esclusivamente indicatori quantitativi senza tener conto di altri elementi come la diffusione della pubblicazione, l'impatto dell'articolo sul sistema produttivo e sociale, il grado di innovazione della ricerca, si rischia di incentivare attività poco produttive o anche inutili.

³¹ Centro universitario per la valutazione e il controllo CRESCO. Report sulla valutazione della ricerca scientifica universitaria. A cura di Maria Vella. Gennaio 2004
<http://www.cresco.unisi.it/wpeng/wp2eng.pdf>

³² Guido Trombetti, Presidente della CRUI Conferenza dei Rettori delle Università Italiane, Relazione sullo stato delle Università Italiane 2006. Roma, 9 novembre 2006
www.cruai.it/cruai/newsletter/ottobre/lettera_CRUI_10.pdf

7. Indici bibliometrici alternativi all'Impact Factor

Il Web ha fatto sorgere centinaia di nuovi database o strumenti che permettono la ricerca di citazioni. Questi strumenti ricadono entro tre categorie. La prima permette all'utente di cercare nel campo a testo pieno per determinare se un certo item, autore o periodico, sono stati citati in un documento.

Rientrano in questa categoria: *arXiv.org* (*eprints server*), *Google Book Search* (*progetto digitalizzazione testi*), *Journal of Archive* di IP, *Physical Review Online Archivi*, *Scirus* di Elsevier (*motore di ricerca*), *CiteSeer* (*automated o autonomus citation indexing ACI*), *Google Scholar* (*motore di ricerca con sistema di ranking basato su ACI*). Il sistema ACI, *Autonomous Citation Index*, è una sorta di motore di ricerca che al suo interno consente l'indicizzazione automatica con estrazione dal testo pieno di parole significative, che vengono contestualizzate e rendono operativo il linking tra le risorse organizzate nelle sue directory. Solitamente è applicato a biblioteche digitali ad accesso aperto.

La seconda categoria di database o strumenti permette all'utente di ricercare nel campo dei riferimenti citati (articoli che citano altri articoli) per identificare citazioni rilevanti. Questi strumenti esistono dalla fine degli anni novanta quando i database disciplinari cominciano ad aggiungere informazioni ai riferimenti citati entro i loro record. Esempi di questi database sono: *MathSciNet*, *Science Direct* di Elsevier, *SciFinder Scholar* di CAS, *SPIRES-HEP* (*eprints server* di Stanford), *Scicitation/SPIN* di AIP, *NASA Astrophysics Data System*.

La terza categoria è composta da database che lavorano esattamente come WoS. Recentemente il principale e forse unico esempio che rientra in questa stessa categoria è *Scopus*, lanciato nel 2004 da Elsevier. Sebbene esso copra molti più periodici referati e atti di conferenze di WoS, fornisce ricerche di citazioni solo dal 1996 in avanti, laddove il WoS arriva ai primi del 1900.

L'impact factor non è comunque più il solo indicatore bibliometrico a cui fare riferimento, numerosi sono gli indici citazionali che segnano una svolta epocale in questo affascinante campo di indagine.³³

Nel mondo Open Access stanno nascendo una serie di iniziative e progetti che interessano il campo della "bibliometria" dove si stanno sperimentando soluzioni alternative al tradizionale IF Impact Factor³⁴.

8. L'Eigenfactor (EI)

Uno di questi metodi è *Eigenfactor*³⁵, un nuovo modello matematico per calcolare l'impatto delle riviste scientifiche accademiche che tiene conto dei diversi standard di citazione tra una disciplina e l'altra. Si tratta di un progetto di ricerca accademica non-commerciale sponsorizzato dal Bergstrom lab al Department of Biology dell'Università di Washington che ha lo scopo di sviluppare un nuovo metodo per la valutazione dell'influenza (impatto) dei periodici accademico-scientifici ai fini di ottenere una mappatura della struttura della ricerca accademica.

I sei punti di forza di Eigenfactor³⁶ sono:

³³ Rossana Morriello, L'indice di Hirsch (h-index) e altri indici citazionali dopo l'impact factor, "Biblioteche oggi", 25 (2007) 1, p. 23-32.

³⁴ A riguardo si veda il lavoro di Marie McVeight della Thomson Scientific (2004) sull'analisi del fattore di impatto e i pattern citazionali dei periodici ad accesso aperto entro i database dell'ISI

<http://www.thomsonscientific.com/media/presentrep/essayspdf/openaccesscitations2.pdf>

³⁵ Eigenfactor: ranking and mapping scientific knowledge

<http://www.eigenfactor.org/>

³⁶ <http://www.eigenfactor.org/whyeeigenfactor.htm>

1. Effettua il posizionamento (rank) dei periodici nello stesso modo in cui Google effettua il ranking per i siti web, proprio perché usa le strutture dell'intera rete (piuttosto che informazioni citazionali puramente localizzate) per valutare l'importanza di ciascun periodico
2. Misura oltre l'influenza della citazione anche il prezzo dei periodici in collaborazione con il sito *Journal Cost-Effectiveness*³⁷, fornendo informazioni sul prezzo e sul valore di migliaia di periodici scientifici. Mentre le funzionalità in ricerca per indicatore bibliometrico EF (Eigenfactor) o per indicatore AI (Article Influence) non comprendono direttamente nel record l'informazione del prezzo, la funzionalità offerta dalla maschera di ricerca *Cost-Effectiveness Search*³⁸ offre informazioni assai interessanti in relazione al fattore costi-benefici
3. Contiene 115.000 reference items, non solo relativi ad articoli pubblicati su oltre 7000 periodici, ma anche tesi di dottorato, e provenienti anche dai campi delle scienze sociali
4. Equilibra le differenze citazionali tra discipline dovute a differenti standard citazionali o differenti scale temporali entro le quali le citazioni avvengono, permettendo una migliore comparazione tra aree di diverse aree di ricerca.
5. Usa i dati delle citazioni su base quinquennale, ciò perché in certe discipline due anni potrebbero essere insufficienti
6. E' completamente free e interamente ricercabile.

9. L'Indice di Hirsch o H-index e sue varianti

Da qualche anno è nato anche un altro "indice", l'**H-index o Indice di Hirsch** noto anche come **indice H**³⁹, che si basa sul numero delle pubblicazioni di un autore ed sul numero di citazioni ricevute. Fu proposto nel 2005 da Jorge E. Hirsch della University of California di San Diego per quantificare la prolificità e l'impatto del lavoro degli scienziati.

Nasce e funziona in modo diverso dall'Impact Factor.

Hirsch spiega nel suo lavoro - apparso sui prestigiosi *PNAS Proceedings of the National Academy of Sciences*⁴⁰ - che

uno scienziato possiede un indice H se H dei suoi N_p lavori hanno almeno H citazioni ciascuno e i rimanenti (N_p - H) lavori hanno ognuno meno di H citazioni.

Sappiamo bene che le convenzioni riguardo le pubblicazioni variano entro differenti comunità che hanno regole diverse in relazione alla disciplina: in fisica e in matematica, un ricercatore che è moderatamente produttivo possiede un H pari al numero di anni di lavoro, mentre scienziati che operano nel campo medico o biologico tendono possedere valori più elevati. Per esempio un matematico solitamente ha un Indice H tra il 4 -5.

³⁷ *Journal Cost-Effectiveness* (2005-6 BETA)

<http://www.journalprices.com/>

³⁸ <http://www.eigenfactor.org/pricereach.php>

³⁹ Sono disponibili su web programmi per il calcolo dell'h-index

<http://www.epidemiologic.org/2006/12/h-index-calculator-of-scientist-impact.html>

⁴⁰ Il suo lavoro lo si trova proprio nell'archivio aperto aXiv dei fisici

<http://arxiv.org/abs/physics/0508025>

dove si può vedere anche il Trackbacks relativo

<http://arxiv.org/tb/physics/0508025>

ArXiv usa il Trackbacks da qualche anno, forse dal 2002, si tratta di un meccanismo mutuato dal Blog per lo scambio comunicativo.

Hirsch inoltre ha dimostrato che H ha un alto valore di previsione sia che uno scienziato abbia vinto o meno premi come il Nobel o sia membro di qualche importante accademia.

L'interesse che ha suscitato l'indice H rispetto ad altri meccanismi è che si può calcolare in modo veloce con i semplici database di Internet, ma soprattutto perchè verifica la reale influenza di uno scienziato sulla comunità, prescindendo da singoli articoli di grande successo, o anche dai lavori di autori che pur avendo pubblicato molto, hanno prodotto solo articoli di scarso interesse, come invece avviene usando l'IF. La sua efficacia si dimostra soprattutto per un confronto tra scienziati che lavorano nello stesso campo, in particolare i fisici e i matematici.

Per esempio l'indice Hirsch di Stephen Hawking è pari a 62, indice basato su SPIRES il database di Stanford, connesso agli archivi del CERN⁴¹.

Il calcolo dell'indice H viene eseguito in base alla distribuzione delle citazioni che le pubblicazioni di un ricercatore ricevono, sebbene lo si possa applicare anche a piccoli gruppi di ricerca entro un ateneo. Il sistema è piuttosto semplice. Dato un autore (o un gruppo di autori, se si vuole applicare tale indice ad un gruppo di ricerca) si procede come segue:

1. Si definisce un certo arco temporale, ad esempio dal 2001 ad oggi.
2. Si ordinano le pubblicazioni in senso decrescente di citazioni, mettendo in testa il lavoro di quell'autore che ha ricevuto un maggior numero di citazioni
3. L'indice H è identificato dal numero sequenziale della pubblicazione che precede quella ove il numero di citazioni risulta inferiore al numero sequenziale della pubblicazione

Vediamo praticamente come funziona il calcolo dell'Indice H usando il calcolatore ***h-Index Calculation and Charts on QuadSearch: Metasearch Engine***⁴² su uno specifico autore.

Scientific Search. Displaying 1-145 of 145 results for query: de robbio (3.41 sec)

H-INDEX: 5

Maximum Cites: 9

Total Cites: 116, Total Articles: 145

Cites/Paper: 0.8

Tabella 1. Esempio di calcolo dell'Indice di Hirsch per autore "De Robbio" con ***h-Index Calculation and Charts on QuadSearch: Metasearch Engine***

⁴¹ <http://www.slac.stanford.edu/spires/>

⁴² <http://quadsearch.csd.auth.gr/index.php?lan=1&s=2>

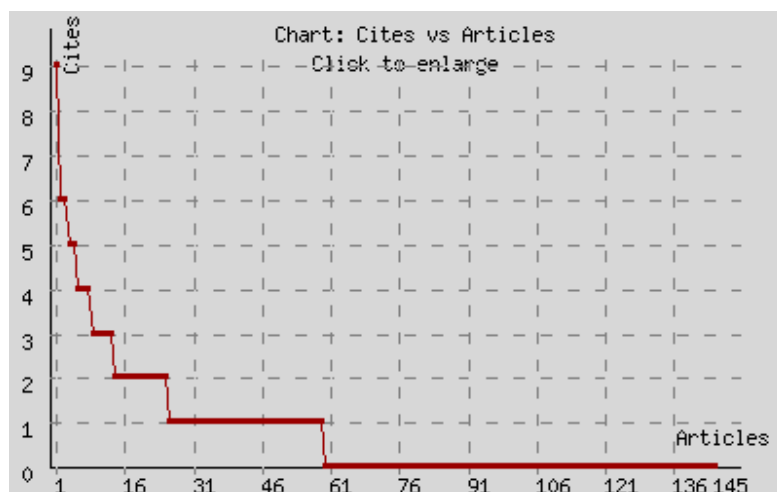


Grafico 1. : Autore “De Robbio” 145 articoli, citazioni totali ricevute 116, max 9 citazioni per articolo

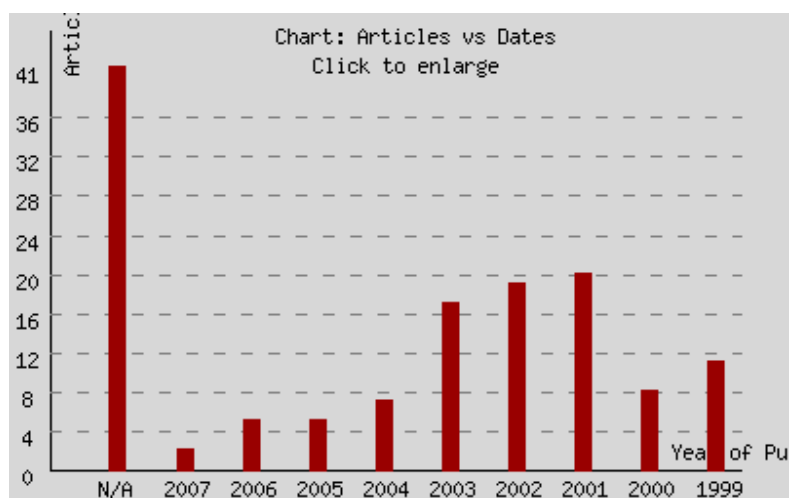


Grafico 2. : Autore “De Robbio”i 145 articoli, suddivisi per anno di pubblicazione

1. ☐ [E-LIS: an international open archive towards building open digital libraries](#) Calculate
A De Robbio, I Subirats Coll - High Energy Physics Libraries Webzine (11/2005), 2005
Cited by 9
Google Scholar Rank: 1
2. ☐ [Il Sistema Bibliotecario di Ateneo e lo sviluppo del mercato dei periodici elettronici.](#) Calculate
A Opocher, A De Robbio, E Marinoni, R Negriolli, R ... - cab.unipd.it
Cited by 6
Google Scholar Rank: 27
3. ☐ [La tutela giuridica delle banche nel diritto d'autore e nei diritti connessi.](#) Calculate
A De Robbio - 1999 - eprints.rclis.org

Cited by 6
Google Scholar Rank: 108

4. ☐ MAI: MetaOPAC Azalai Italien

Calculate

A **De Robbio** - La Revue BIBLIOthèque (s, 2002
Cited by 5
Google Scholar Rank: 31

5. ☒ [Periodici elettronici nel ciberspazio \[Italian\]](#)

Calculate

A **De Robbio** - eprints.rclis.org
Cited by 5
Google Scholar Rank: 39

6. ☐ [Mathematics Subject Classification and related schemes in the OAI framework. In: Find and Post ...](#)

Calculate

A **De Robbio**, D Maguolo, A Marini - 2002 - eprints.rclis.org
Cited by 4
Google Scholar Rank: 17

7. ☐ [Mathematics Subject Classification \(MSC\) and related Schemes in the Digital World](#)

Calculate

A **De Robbio**, D Maguolo, A Marini - 2001 - eprints.rclis.org
Cited by 4
Google Scholar Rank: 24

8. ☐ [Impacts de la directive européenne sur le droit d'auteur en Italie](#)

Calculate

A **De Robbio** - Bulletin des bibliothèques de France(Imprimé), 2006 - cat.inist.fr
Cited by 4
Google Scholar Rank: 34

Tabella 2.: L'Indice Hirsch per autore "De Robbio" è pari a 5. La prima pubblicazione che si incontra nella lista sequenziale in ordine decrescente con un numero di citazioni più basso del suo numero di sequenziale è collocata in sesta posizione. Pertanto la precedente pubblicazione ha numero di sequenza pari a 5, con 5 citazioni.

L'indice H è tuttavia meno ingenuo di quanto potrebbe sembrare a prima vista. Infatti è stato usato, o meglio studiato in modo approfondito, da alcune comunità dove sarebbe risultato penalizzante in quanto discutibile per certi aspetti. Si sono presi a modello due matematici tra i più influenti e precisamente Louis Nirenber e Peter D. Lax ed è stato calcolato l'indice di Hirsch che è risultato pari a 39 e 33 rispettivamente. L'indice è anche influenzato dal contesto in cui viene calcolato, cioè "su" quale database o quali insiemi di database. Gli studi presenti in letteratura solitamente sono stati condotti sui data base dell'ISI proprio per ragioni di comparazione dei due metodi (Indice H e IF). Ovviamente l'indice H ha anche dei limiti per scienziati che hanno carriera breve, vedi Galois morto in duello all'età di 20 anni che ha un indice pari a 2 o lo stesso Einstein al quale l'indice H non rende di certo onore... ;-)

Publish or Perish⁴³ implementa anche un metodo alternativo individuale sul modello H-index con un differente approccio: invece di dividere il totale H-index, prima normalizza il numero di citazioni per ciascuno dei lavori attraverso la divisione del numero di citazioni per il numero di autori per quel lavoro, poi calcola l'indice H del conteggio di citazione normalizzato.

Egghe e Rousseau⁴⁴ in un interessante lavoro del 2006 propongono un modello “infometrico” per l'applicazione dell'indice H.

Egghe peraltro è l'autore del metodo alternativo all'H-index, noto come G-index. L'indice G o **g-index** è un indicatore bibliometrico di tipo quantitativo per misurare la produttività scientifica dei fisici e altri scienziati, anch'esso basato sui record citazionali delle loro pubblicazioni. Fu suggerito nel 2006 da Leo Egghe, sulla base di una variante dell'indice H. Il g-index è calcolato sulla base della distribuzione delle citazioni ricevute dalle pubblicazioni di un dato ricercatore.

Given a set of articles ranked in decreasing order of the number of citations that they received, the g-index is the (unique) largest number such that the top g articles received (together) at least g^2 citations.

Mentre Antonis Sidiropoulos, Dimitrios Katsaros, e Yannis Manolopoulos⁴⁵ hanno descritto come migliorare l'indice H dando più peso ai lavori recenti "*thus rewarding academics who maintain a steady level of activity*", Pablo D. Batista, Monica G. Campiteli, Osame Kinouchi, e Alexandre S. Martinez⁴⁶ propongono di dividere l'indice H standard per il numero medio di autori nell'articolo che contribuiscono all'indice stesso, di modo da ridurre gli effetti del co-authorship.

Altra variante dell'indice H è l' **h-b-index (Hirsch-Banks Index)**, meglio un'estensione dell'indice H sviluppato da Michael Banks del Max Planck Institute for Solid State Research, focalizzato sulle pubblicazioni in fisica dello stato solido. E' definito allo stesso modo dell'Indice H, ma è basato sulla ricerca di uno specifico argomento o composto, invece che del nome dello scienziato.

Nel caso di un argomento è utile definire l'h-b index in termini di numero di anni, n come $h = nm$

Se l'h-b index è lineare con il numero di anni, allora m è dato come il gradiente. A questo riguardo, un composto o argomento specifico con un alto fattore m e h-b index può essere definito come un *hot topic*.

Il curatore di *Science Focus*, Jin Bihui ha inventato un ulteriore metodo, detto “**a-index**” il quale è definito come la media del numero di citazioni ricevute da opere nel numero di pubblicazioni “h-index”.

Lo stesso Rousseau⁴⁷, in uno dei suoi lavori depositati in E-LIS⁴⁸, dice che è dimostrato che l'h-index su una mano e l'a-index e g-index sull'altra, misurano cose differenti. L'a-index tuttavia sembra essere

⁴³ Vedere per esempio il programmino sul sito Citation metrics di Harzing

<http://www.harzing.com/resources.htm#/pop.htm>

⁴⁴ Egghe, L.; Rousseau R. (2006) An informetric model for the Hirsch-index, *Scientometrics*, vol. 69, no. 1, pp. 121-129.

⁴⁵ Antonis Sidiropoulos, Dimitrios Katsaros, and Yannis Manolopoulos, Generalized h-index for disclosing latent facts in citation networks, arXiv:cs.DL/0607066 v1 13 Jul 2006,

⁴⁶ Pablo D. Batista, Monica G. Campiteli, Osame Kinouchi, and Alexandre S. Martinez, Is it possible to compare researchers with different scientific interests?, *Scientometrics*, Vol 68, No. 1 (2006), pp. 179-189.

⁴⁷ Rousseau, Ronald (2006) New developments related to the Hirsch index.

<http://eprints.rclis.org/archive/00006376/>

⁴⁸ E-LIS è l'archivio aperto internazionale di ambito LIS *Library and Information Sciences*

<http://eprints.rclis.org/>

eccessivamente sensibile ad un articolo estremamente e altamente citato. Per questo motivo sembrerebbe il g-index sia il più utile dei due. Quanto all'h-index e al g-index essi misurano differenti aspetti di una lista di pubblicazioni di uno scienziato. Sebbene il g-index sia un indicatore più sensibile dell'h-index, nessuno dei due ne racconta la storia completa. Presi assieme g e h index presentano un'immagine concisa dei successi di uno scienziato in termini di pubblicazioni e citazioni. In un altro dei suoi lavori Rousseau⁴⁹ illustra l'evoluzione dell'Hirsch index applicato ad un caso di studio sul Journal of the American Society of Information Science nel periodo 1991-2000.

10. La metrica del Web e il Web Impact Factor (WIF)

Sebbene l'analisi delle citazioni non sia nulla di nuovo ad oggi la potenza dei grandi calcolatori consente un'applicazione molto più estesa. Il sistema di ranking di motori di ricerca come *Google's PageRank*⁵⁰ è basato sul principio dell'analisi citazionale. Si parla quindi di **metriche del Web** o **Webometrica**. La scienza detta **Webometrics** o **Webmetrics** o anche **cybermetrics**, o **web metrics** (metrica del web) tenta di misurare il World Wide Web al fine di ottenere conoscenze sul numero e sul tipo di connessioni ipertestuali gli hyperlinks, sulle strutture del Web e sui modelli di utilizzo delle risorse che popolano il cyberspazio. De Bellis, nella sua tesi pubblicata su Biblioteche Oggi⁵¹ traccia un percorso assai interessante che ripercorre a ritroso il contesto della bibliografia statistica, a partire dalle leggi che definisce "leggi universali" utili a "contare la scienza", come la legge di Lotka e la legge di Bradford, poco conosciute in Italia e anche in Europa, ma molto note ai bibliotecari statunitensi a seguito della crisi del 1929 in quanto usate in applicazione agli acquisti di periodici scientifici in funzione della costruzione delle raccolte. De Bellis contestualizza la citazione bibliografica, come dice il titolo stesso della sua tesi, nell'epoca della sua riproducibilità tecnica, analizzando mappe e paradigmi entro un assetto storico e di sociologia della scienza e il Fattore di Impatto e gli strumenti cosiddetti ISI entro un assetto di politica della scienza, per confluire ai modelli attuali della Cybermetrica.

Il PageRank è un termine ormai entrato nel lessico quotidiano, grazie ai servizi offerti dal motore di ricerca Google. Si riconduce al concetto di popolarità tipico delle relazioni sociali umane, e sta a significare le pagine o i siti di maggiore rilevanza in relazione ai termini ricercati. Il sistema usa algoritmi⁵² basati su indicizzazione del materiale raccolto da Google e sui collegamenti presenti nelle pagine esaminate, oltre che sulla pertinenza del contenuto di una pagina, nonché sul numero di pagine correlate, in risposta ai termini ed ai criteri della ricerca effettuata. Questo meccanismo permette di attuare un minimo di controllo incrociato che garantisca la validità dei risultati di ricerca. Maggiore è il grado di popolarità di un sito, maggiore è il valore dei voti (*link*) che quello stesso sito può esprimere. Sebbene Almind e Ingwersen nel 1997 coniarono per primi il termine **webometrics**, fu solo nel 2004 che Björneborn e Ingwersen ne tracciarono la definizione come "*lo studio delle funzioni quantitative*

Il sistema di classificazione JITA di E-LIS evidenzia nella categoria BA e BB numerosi lavori di ambito bibliometrico e di uso e impatto dell'informazione

BA. Use and impact of information. (197)

BB. Bibliometric methods. (241)

⁴⁹ Rousseau, Ronald (2006) A case study : evolution of JASIS' Hirsch index.

<http://eprints.rclis.org/archive/00005430/>

⁵⁰ http://www.google.it/why_use.html

⁵¹ De Bellis, Nicola [2005] La citazione bibliografica nell'epoca della sua riproducibilità tecnica: bibliometria e analisi delle citazioni dallo Science Citation Index alla Cybermetrica, ultima revisione 31/05/2005,

<<http://www.bibliotecheoggi.it/content/CITAZIONE.pdf>>

⁵² L'algoritmo completo per il calcolo del PageRank fa ricorso all'uso della teoria dei processi di Markov

della costruzione e dell'uso delle risorse informative, delle strutture e tecnologie nel del Web disegnato sugli approcci bibliometrici e infometrici⁵³”.

Una misura relativamente diretta è il **Web Impact Factor (WIF)** definito nel 1998 da Ingrwersen come il numero di pagine totali di collegamento, diviso dal numero di Web pagine pubblicate nei siti che sono accessibili ai motori di ricerca. Il collega iraniano Alireza Noruzi⁵⁴, che ha prodotto una serie di interessanti lavori sul Web Impact Factor, ha rilevato che il WIF funziona in modo adeguato soltanto all'interno della webosfera di un singolo Paese, usando una determinata lingua e un argomento definito. Interessanti sono i risultati emersi dal suo studio sul fattore di impatto web entro le università in Iran.⁵⁵

11. Iniziative per la standardizzazione delle statistiche e per la misurazione dei dati di utilizzo

Se gli indicatori generati dall'autore sono le citazioni, quelli generati dai lettori sono i dati sull'utilizzo. Questi dati possono essere raccolti tramite server web e *linkresolver logs*. E' necessario però normalizzare il lavoro di raccolta dei dati di log al fine della condivisione di metodi per ottenere analisi significative.

Sebbene lo sviluppo di indicatori quantitativi ai fini dell'impatto della ricerca basati sui dati d'utilizzo degli utenti sia ancora all'inizio, i dati d'uso sono da sempre considerati come una risorsa di un certo valore nella valutazione degli oggetti della ricerca scientifica. Le citazioni possono essere analizzate alla stregua dei dati di utilizzo, quantitativamente o strutturalmente. Queste analisi possono combinarsi tra loro o completarsi con altri indicatori per creare nuove misurazioni alternative da aggiungere al fattore di impatto.

John Ewing, dell'American Mathematical Society in un suo articolo del 2006⁵⁶ dice che è tempo di smetterla di dare risposte semplicistiche a complicate questioni di giudizio. L'Impact Factor sicuramente non è inutile, ma misurare il valore di una rivista attraverso in indicatore di impatto non è la stessa cosa che misurare il valore di un articolo o di autore nella sua interezza: la misurazione della qualità richiede molte più informazioni rispetto a quelle che un indicatore come l'IF può dare. Nemmeno le statistiche di utilizzo sono inutili, prosegue Ewing, ma da sole non possono misurare il valore di un articolo scientifico, anzi se le statistiche di utilizzo possono essere leggermente utili, la loro misurazione può essere enormemente dannosa. Il giudizio scientifico, a detta di Ewing, è il miglior indicatore per una corretta valutazione, sempre ammesso che non sia troppo soggettivo o che non vi sia conflitto di interesse, aggiungono altri autori a commento dell'articolo di Ewing.

In sostanza Ewing sembra accettare la tradizione di un IF sebbene imperfetto e non equo - infatti cita degli esempi abbastanza eclatanti in termini di enormi scarti di valore IF tra riviste di medicina e di matematica – ma sembra non accettare affatto l'idea di applicare indicatori di utilizzo delle risorse, forse perché concetto ancora non del tutto maturo.

⁵³ Leo Egghe & Ronald Rousseau, *Introduction to Informetrics: Quantitative Methods in Library, Documentation and Information Science*. Elsevier, 1990. 978-0444884930

⁵⁴ Noruzi, Alireza (2006) *The Web Impact Factor : a critical review*. *The Electronic Library* 24(4):pp. 490-500.
<http://eprints.rclis.org/archive/00005543/>

⁵⁵ Noruzi, Alireza (2005) [Web impact factors for Iranian Universities](http://eprints.rclis.org/archive/00005596/). *Webology* 2(1).
<http://eprints.rclis.org/archive/00005596/>

⁵⁶ John Ewing *Measuring Journals*
Notices of the AMS, Vol. 53, No. 9, October 2006, 1049-1053
<http://www.ams.org/notices/200609/comm-ewing.pdf>

Gli archivi aperti ben si prestano ad analisi bibliometriche in quanto si tratta di campi di indagine ben definiti che possono condurre ad una valutazione dell'impatto nella ricerca non tanto delle riviste, bensì del singolo autore o meglio del singolo lavoro.

Al recente workshop OAI5 dello scorso aprile 2007⁵⁷ alcuni interventi si sono focalizzati proprio sulla misurazione delle risorse entro gli archivi aperti e risorse Open Access. Johan Bollen del Digital Library Research & Prototyping Team presso la biblioteca del Los Alamos National Laboratory Research ha presentato il programma *MESUR*⁵⁸ *metrics from scholarly usage of resources*, che offre un set di strumenti per la valutazione degli oggetti della comunicazione scientifica. Il progetto biennale MESUR, finanziato dalla Andrew W. Mellon Foundation, mira a definire e rendere valide una serie di misure basate sull'impatto dei dati di utilizzo, e a formulare linee guida nel rispetto delle caratteristiche e delle proprie applicazioni.

Il progetto MESUR sta costruendo un modello semantico su larga scala ad utilizzo della comunità scientifica, che sembra integrare bene un gran numero di dati bibliografici, citazionali e di utilizzo. Funzionando come un set di dati di riferimento, questo modello è analizzato per caratterizzare l'intricata rete di relazioni che esiste nella comunità scientifica. Il risultato è un programma per la definizione e la validazione delle misure basate sull'impatto d'uso.

L'impatto di una pubblicazione infatti può essere misurato da indicatori generati da parte dall'autore, con lo strumento della citazione bibliografica che, come abbiamo visto è uno strumento collaudato. Ma è anche misurabile sull'altro versante, cioè da indicatori generati dal lettore che indicano quanto una risorsa è fruita. Entro il circuito della comunicazione scientifica, il lettore è egli stesso autore di lavori analoghi, e comunque membro appartenente alla comunità o allo stesso campo di ricerca.

Al citato workshop OAI5 Leslie Carr ha sottolineato invece come ad oggi l'uso dei dati di utilizzo dei repository e dei loro contenuti sia potenzialmente sottovalutato, ma come vedremo più avanti i dati di utilizzo delle risorse sono misurazioni che nascono proprio dal contesto commerciale e non tanto dal mondo OA. Una comprensione dell'utilizzo di tali dati è altamente desiderabile sia per gli autori stessi sia per i responsabili del deposito (istituzioni e loro amministratori), oltre che per chi deve valutare la ricerca ai fini della distribuzione dei finanziamenti. Un certo numero di iniziative localizzate nei singoli archivi hanno elaborato statistiche di vario genere, ma la sfida reale sarà quella di elaborare statistiche che possono essere raccolte e confrontate in modo trasparente su scala globale, anche in un confronto aperto con i sistemi usati da editori commerciali. Esistono iniziative volte a fornire infrastrutture adeguate che includono i dati dell'editore e i dati raccolti attraverso gli archivi ad accesso aperto: *COUNTER* (*Counting Online Usage of NeTworked Electronic Resources*), *SUSHI* (*Standardized Usage Statistics Harvesting Initiative*), *IRS* (*Interoperabile Repository Statistics*, *CiteBase*⁵⁹, etc...

COUNTER (*Counting Online Usage of NeTworked Electronic Resources*)⁶⁰ è un'iniziativa detta multi-agency, attivamente supportato dalle comunità internazionali di bibliotecari ed editori oltre che

⁵⁷ 5th Workshop on Innovations in Scholarly Communication. CERN, Geneva, Switzerland, 18-21 aprile 2007

<http://oai5.web.cern.ch/oai5/>

⁵⁸ <http://www.mesur.org/Home.html>

Relazione adOAI5

<http://indico.cern.ch/getFile.py/access?contribId=6&sessionId=14&resId=1&materialId=slides&confId=5710>

⁵⁹ Maria Cassella, Novità recenti dal mondo delle statistiche di uso: il protocollo SUSHI e le nuove linee guida dell'ICOLC. AIB notizie 18 (2006), n. 9

<http://www.aib.it/aib/editoria/n18/0908.htm3>

⁶⁰ La prima realease fu pubblicata già nel gennaio 2003 ed ora è ampiamente implementata. La release 2. pubblicata nell'aprile 2005 dal gennaio 2006 ha definitivamente rimpiazzato la prima versione.

www.projectCounter.org

dalla organizzazioni professionali, ha l'obiettivo di sviluppare un set estensibile e accettato a livello internazionale di codici di pratiche che permettono la misurazione di utilizzo di prodotti e servizi informativi online in un modo credibile, consistente e compatibile.

Lo standard COUNTER comunque non offre da solo una risposta esaustiva ed adeguata, in quanto pur fornendo un ottimo modello per la produzione di statistiche d'uso, non soddisfa totalmente i bisogni della comunità. Ciò è stato posto in evidenza in particolare dai bibliotecari e da chi deve "leggere" e "interpretare" i rapporti di COUNTER che sono generati su fogli Excel, con dati poco aggregabili elaborati da ciascun editore e suddivisi in file differenti.

Si rendeva necessario un modello di scambio di dati più efficiente. L'iniziativa *SUSHI, Standardized Usage Statistics Harvesting Initiative*, che risale al novembre del 2004 sembra ad oggi essere la soluzione più naturale proprio perché usa un protocollo basato sui servizi web e sugli schemi XML. Nasce come diretta conseguenza dei primi utilizzi di software per la gestione delle risorse elettroniche detti ERM Electronic Resource Managers systems, gestione che deve coprire tutte le fasi del ciclo di vita di una risorsa: acquisizione, trial se esistente, costi e relativo monitoraggio, pagamenti, modalità di accesso, dati statistici. Sebbene allo stato attuale SUSHI non sia un modello che permette calcoli su statistiche d'uso, i dati che SUSHI acquisisce sono sempre nel formato COUNTER, in quanto vengono conservati gli schemi XML⁶¹.

12.L'Usage Factor (UF)

Quanto sopra è riconducibile ad una nuova strada o modello di misurazione, definibile come un indicatore bibliometrico quantitativo di nuova generazione, l'*Usage Factor UF*, complementare alla tradizionale modello noto dell'*Impact Factor IF*.

L'UK Serials Group (UKSG) ha commissionato ricerche entro un progetto che ha lo scopo di sondare la possibilità di utilizzo di un Usage Factor come indice per misurare il valore e l'autorevolezza delle riviste scientifiche e pubblicazioni accademiche in generale.

La prima fase del progetto si è conclusa nel febbraio 2007 e ha raccolto una serie di interviste approfondite ai principali attori del mercato della comunicazione scientifica⁶². La seconda fase consisterà in una serie di indagini rivolte sia a bibliotecari sia agli autori e somministrate via web con l'obiettivo di comprendere meglio alcuni punti emersi durante la prima fase e di ottenere risposte su alcuni punti chiave verificando le conclusioni raggiunte nel corso della prima fase.

In sostanza il progetto mira a sondare la fattibilità reale di uno sviluppo di un nuovo indice bibliometrico di misurazione del valore dei periodici scientifici basato sull'utilizzo delle risorse elettroniche che potrebbe essere appunto definito come Usage Factor o Fattore di Utilizzo.

Su questa linea un recente studio sponsorizzato dal britannico JISC ha sottolineato alcune considerazioni⁶³:

- L'Impact Factor dell'ISI, basato sull'analisi citazionale, è divenuto uno strumento generalmente accettato come valida misurazione della qualità dei periodici scientifici, ed è ampiamente usato

⁶¹ Per un lavoro più approfondito sul modello SUSHI si veda

D. Dellisanti, ICOLC 2006: SUSHI un nuovo strumento per la misurazione delle risorse elettroniche, Bollettino CILEA Nr. 105 Febbraio (2007)

http://bollettino.cilea.it/rst/rst.php?op=view_metadata&id=549

⁶² Le interviste e i risultati della seconda fase sono stati oggetto di discussione durante la trentesima UKSG Annual Conference che si è svolta dal 16 al 18 aprile 2007 scorso presso l'Università di Warwick.

⁶³ http://www.jisc.ac.uk/uploaded_documents/nesli2_usstudy.pdf

da editori, autori, agenzie di finanziamento e bibliotecari come misura della qualità di un periodico.

- ❑ Cionostante ci sono numerosi dubbi sull'uso di tale fattore come indice bibliometrico considerate le critiche mosse da alcune comunità
- ❑ La disponibilità online della quasi totalità dei periodici scientifici significativi, combinata con la disponibilità di una sempre maggiore credibilità di statistiche di utilizzo prodotte online da sistemi COUNTER compatibili, aumenta la possibilità di una misurazione parallela della performance del periodico basata non solo sull'IF ma anche sull'utilizzo della risorsa. Ciò diviene un percorribile indice metrico addizionale
- ❑ Tale indice bibliometrico definito come UF "Usage Factor potrebbe essere basato sui dati contenuti nelle specifiche di COUNTER Journal Report 1

$$\text{Usage Factor} = \frac{\text{Total usage (COUNTER JR1 data for a specified period)}}{\text{Total number of articles available online (for the same specified period)}}$$

È riconosciuto che, a differenza dei dati di citazione, le statistiche di utilizzo totali globali COUNTER-compliant per le diverse pubblicazioni, attualmente non sono pubblicate, e non lo saranno nel prossimo futuro. Per gli scopi di questo studio, quindi, sarà necessario derivare l'UF dai dati di utilizzo detenuti da una selezione di biblioteche. L'obiettivo generale dello studio sarà determinare se il concetto di UF sia davvero significativo, se sarà pratico da implementare e se fornirà adeguati chiarimenti e un osservatorio atto a stabilire un corretto valore alla misura della qualità delle pubblicazioni online.

13. Le nuove frontiere delle analisi bibliometriche

Di seguito si vogliono passare in rassegna alcuni esempi di nuovi strumenti e/ progetti per l'analisi citazionale e per l'analisi statistica dei log e dei dati⁶⁴, solo a titolo di breve informazione.

IRS: Interoperable Repository Statistics ⁶⁵ è un progetto entro il JISC Digital Repositories proposto da University of Southampton e Key Perspectives Ltd e supportato da University of Tasmania, Australia, Long Island University, USA, e il progetto COUNTER, finora usato solo da editori commerciali e perciò su lavori che sono stati sottoposti a peer-review. Su questa base, entro il framework OAI, un Service Provider potrà raccogliere e aggregare dati di utilizzo da archivi istituzionale o altre fonti, come archivi disciplinari e periodici ad accesso aperto OA. Il progetto costruirà un pilot per la dimostrazione delle analisi statistiche dei service provider, utile a comparare tutti i dati statistici delle varie versioni di un lavoro sia pubblicato su riviste tradizionali sia invece depositato su un archivio aperto oppure i dati di utilizzo di riviste ad accesso aperto in comparazione con quelli di riviste ad accesso chiuso.

⁶⁴ Buttà Basilio, *Metadati negli Open Archive: tecniche di conversione dei formati dei record nel formato MARC21 utilizzato da CDSware*, tesi di laurea in informatica, Università degli Studi di Messina, A.A. 2003-2004 (relatore Puccio L., correlatore De Robbio A.).

Amelotti Ercole, *Protocollo OAI-PMH negli Open Archive e applicazione CDSware per la rappresentazione dei relativi dati bibliografici*, tesi di laurea in informatica, Università degli Studi di Messina, A.A. 2003-2004 (relatore Puccio L., correlatore De Robbio A.).

⁶⁵ IRS (Interoperable Repository Statistics)
<http://irs.eprints.org/about.html>

CiteBase⁶⁶ è uno strumento di analisi citazionale relativo ai lavori scientifici disponibili on-line. Esso raccoglie pre e post-print da alcuni archivi che si basano sul protocollo OAI-PMH, analizza e collega i loro riferimenti ed indicizza i metadati nel proprio motore di ricerca. La ricerca viene effettuata attraverso tre criteri differenti:

1. metadati: autore/i, titolo/parole chiave, data di pubblicazione/creazione
2. citazioni
3. identificatori OAI

L'applicazione permette all'utente di ottenere, a seguito di ricerche basate sui precedenti criteri, l'organizzazione delle informazioni citazionali in funzione di:

1. numero citazioni dell'articolo
2. numero download articolo
3. numero citazione all'autore
4. numero download lavori dell'autore

Altro strumento fornito da Citebase, di ulteriore valore aggiunto, è il *Generatore di correlazione* che permette di ottenere il grafico della correlazione tra impatto di download e impatto citazionale di un articolo. Attualmente l'applicazione è in fase di sperimentazione presso l'University of Southampton.

CiteSeer⁶⁷, sviluppato presso la Penn State's School of Information Sciences and Technology, è uno strumento di analisi citazionale prevalentemente nel campo dell'informatica, che permette l'indicizzazione delle risorse attraverso citazioni e full-text.

La più importante funzionalità di CiteSeer è l'ACI (Autonomous Citation Indexing), sistema che permette di creare in maniera totalmente automatica un indice basato sulle citazioni per gli scopi di ricerca e valutazione dei lavori.

Altre funzionalità presenti sono:

- Reference linking attraverso link citazionali
- Citation contest per conoscere il grado di interesse di un determinato articolo da parte degli altri ricercatori
- Notifica automatica di nuove citazioni per un dato articolo e di nuovi articoli di interesse

Indicizzazione dei full-text di articoli e delle citazioni

LogEC⁶⁸ è uno strumento atto a collezionare statistiche di accesso da diversi server che utilizzano il set di dati di RePEc. E' quindi uno prodotto per l'analisi statistica degli accessi alle risorse, che consente la valutazione dell'impatto di un determinato articolo.

Tra le funzionalità più importanti si possono considerare:

- aggiornamento mensile delle statistiche a partire dai log raccolti dai server dei siti contenenti gli articoli collezionati
- produzione statistiche significative considerando il numero effettivo di utenti che visualizzano l'abstract o scaricano il fulltext, non considerando gli accessi da parte di spider o robot
- esclusione del conteggio errato dovuto al "double counting"⁶⁹

Awstats⁷⁰ è un analizzatore di log open source per la produzione di statistiche dettagliate e complete per una svariata tipologia di log, da quelli dei Web Server a quelli dei server di e-mail o FTP.

⁶⁶ <http://www.citebase.org/>

⁶⁷ <http://citeseer.ist.psu.edu/>

⁶⁸ <http://logec.repec.org/>

⁶⁹ il termine in questo caso sta a significare il possibile errore in una transazione dove si conteggiano più volte gli stessi dati

⁷⁰ <http://awstats.sourceforge.net/>

Le funzionalità principali di Awstats sono:

- analisi e pubblicazione di statistiche
- analisi dei log suddividendoli per singola unità (es. regione, paese, ecc.)
- distinzione degli accessi dei motori di ricerca da quelli degli utenti

Tra i software open source disponibili a questo scopo, Awstats è sicuramente tra i più completi e attivi.

Paracite⁷¹ è uno strumento di reference linking per la localizzazione di articoli on line a partire dai riferimenti bibliografici. Permette la costruzione del link ad una risorsa a partire dal riferimento bibliografico.

L'applicazione consente di effettuare ricerche a partire dal reference, presentando i risultati ottenuti in base alla priorità dovuta alla maggiore probabilità di poter recuperare l'articolo secondo certi criteri:

- fulltext gratuiti
- qualità del documento

Una delle applicazioni più riuscite è quella usata dall'archivio aperto internazionale E-LIS⁷², dove il codice è stato praticamente riscritto⁷³ ai fini di ottenere parsing più mirati per il recupero tramite il reference linking dei lavori citati.

Altri progetti per stabilire metriche di utilizzo entro un contesto di biblioteca digitale sono: *the ARL E-metrics project*⁷⁴ e *JISC NESLI 2 usage stats project*, condotto dalle Università UCE e Cranfield⁷⁵.

14. Impatto citazionale e usi di fonti informative multiple

L'aumento nell'uso di database su web e degli strumenti di accesso alla letteratura scientifica, ma anche la diversificazione delle varie tipologie di documenti⁷⁶ da analizzare ai fini delle citazioni, e le lingue diverse dal dominante inglese, hanno rilevato quanto vitale sia l'uso di fonti di citazione multiple ai fini di una valutazione accurata dell'impatto e della qualità delle opere prodotte dagli scienziati di tutto il mondo (non solo del Paesi occidentali).

Lokman I. Meho e Kiduk Yang della *School of Library and Information Science* dell'Università dell'Indiana di Bloomington hanno condotto diversi studi per dimostrare la necessità di usare fonti citazionali multiple. In uno dei loro studi⁷⁷ sono stati comparati i risultati della copertura citazionale da WoS, Scopus e Google Scholar su un campione di 25 ricercatori altamente pubblicati nel campo delle scienze dell'informazione. E' stato rilevato che Scopus e Google Scholar aumentano il numero delle citazioni di una media del 35% e 160% rispettivamente. Inoltre, e questo aspetto sembra essere più importante, questo incremento varia considerevolmente da un'area di ricerca ad un'altra, laddove ricercatori che operano nel campo della comunicazione mediata dal computer o interazione uomo-

⁷¹ <http://paracite.eprints.org>.

⁷² <http://eprints.rclis.org>

⁷³ Il codice messo a punto da Zeno Tajoli del Cilea è liberamente scaricabile a <http://eprints.rclis.org/softw.html>

⁷⁴ <http://www.arl.org/stats/newmeas/emetrics/>

⁷⁵ <http://www.ebase.uce.ac.uk/projects/NESLI2.htm>

⁷⁶ Libri, capitoli di libri, eprints, tesi e dissertazioni, relazioni a conferenze, rapporti di ricerca, ...

⁷⁷ Yang, Kiduk, and Lokman I. Meho. "Citation Analysis: A Comparison of Google Scholar, Scopus, and Web of Science." *Information Realities: Shaping the Digital Future for All - Proceedings of the Annual Meeting of the American Society for Information Science and Technology (November 2006)*
<http://www.asis.org/Conferences/AM06/papers/185.html>

computer avevano un tasso di citazione più che raddoppiato rispetto a quelli specializzati in bibliometria a analisi citazionale i quali avevano un tasso di citazioni aumentati meno del 25%.

Un'altra scoperta dello studio all'Università dell'Indiana ha messo in evidenza che l'uso di *Scopus* e *Google Scholar* aveva aiutato a stabilire una connessione tra i vari campi disciplinari dalle scienze dell'informazione alle scienze cognitive, dall'informatica all'ingegneria e anche scienze della formazione, tendenza ad una connessione interdisciplinare che non avrebbe mai potuto essere realizzata solo tramite *WoS*. La presenza di strumenti multipli di citazione permette inoltre di generare mappe molto più accurate di visualizzazione della complessa rete della comunicazione scientifica in generale, utili a stabilire connessioni tra autori, dipartimenti, discipline, periodici o Paesi che si citano o si influenzano l'un l'altro.

Meho ha effettuato un semplice test sul volume *Quantum Comparison and Quantum Information* di M. Nielsen e I. Chung rilevando che in *WoS* questo testo era stato citato oltre 2800 volte, in *Scopus* 3150 e in *Google Scholar* 4300 volte, mentre in *Physical Review Online Archive* 150, in *Science Direct* 375, nel *Journal of Archive* di Institute of Physics 290 e nell'archivio aperto dei fisici *arXiv.org* 325 volte. Se ci si fermasse solo al *WoS* si perderebbero tutte le altre citazioni provenienti dalle altre fonti⁷⁸. Naturalmente molto lavoro è necessario ai fini di una ripulitura della sovrapposizioni che, in particolare per ambiti come *Google Scholar* richiede tempo e strumenti dedicati⁷⁹.

15. Open Access e impatto citazionale: da Lawrence a Moed

Numerosi sono i lavori che evidenziano come un lavoro posto online ad accesso aperto sia notevolmente più scaricato, e di conseguenza si crede più letto di un lavoro posto sempre online, ma su piattaforma commerciale ad accesso chiuso. E' altamente prevedibile che un lavoro scaricato ai fini di una fruizione, sia in un prossimo futuro citato da altri articoli e quindi indicatori di log statistici sono sostanzialmente indicatori di prevedibilità verso un fattore di impatto basato da una parte su una reale fruizione, dall'altra su future citazioni.

Alcuni di questi lavori sono liberamente disponibili per una lettura attenta in merito a dati e casi di studio proposti.

Il primo studio per la valutazione degli effetti del cosiddetto Green OA channel⁸⁰ fu condotto nel 2004 su 95.000 pre-print nei campi della fisica delle matematica presenti nell'archivio aperto *arXiv.org*, da Harnad e Brody⁸¹ e gli elementi dei dati esaminati furono inclusi in alcuni studi successivi sempre dello stesso anno. In questi studi furono comparati i conteggi delle citazioni degli articoli pubblicati ad accesso aperto nel server dei fisici e indicizzati in *WoS* di Thompson con gli articoli pubblicati negli stessi anni presi a campione (tra il 1992 e il 2003) nelle stesse riviste ma non presenti ad accesso aperto. L'uso del conteggio delle citazioni per valutare l'impatto degli articoli di ricerca è un metodo ben consolidato, dicono i due autori. Ciononostante l'impatto di citazione di un articolo può essere misurato solo numerosi anni dopo la sua pubblicazione. Il numero di volte che un articolo è scaricato dalla rete può essere un dato immediatamente registrabile ed enumerabile. Un alto numero di scarichi di un articolo è un possibile indice predittivo di citazioni per quell'articolo negli anni futuri. In sostanza

⁷⁸ Meho, Lokman I. "The Rise and Rise of Citation Analysis." *Physics World* 20, no. 1 (January 2007): 32-36.
<http://www.slis.indiana.edu/faculty/meho/PhysicsWorld.pdf>

⁷⁹ vedi op. cit. Bakkalbasi, Nisa, Kathleen Bauer, and Janis Glover, et al (2006)

⁸⁰ Green sta per via verde si riferisce al canale dell'Open Access che prevede il deposito degli articoli entro gli archivi aperti, istituzionali o disciplinari

⁸¹ Stevan Harnad, Tim Brody, Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals
<http://www.dlib.org/dlib/june04/harnad/06harnad.html>

l'impatto di utilizzo Web è un indicatore a breve termine che predice l'impatto di citazione futuro a medio termine.

Allo scopo di analizzare l'effetto dell'Open Access e dello scarico (*hits*) sull'impatto delle citazioni è stata prediposta dal progetto OpCit⁸² una pagina che raccoglie tutti gli studi in una ricca bibliografia.

Gli articoli e gli studi proposti tracciano un interessante percorso dallo studio originale di Lawrence apparso su *Nature* nel 2001⁸³ e fino ai più recenti lavori di Moed ed altri.

Lawrence constatò nei suoi studi su venti pubblicazioni di punta che il tasso medio di citazione registra un incremento del 286% (mediana 284%) per gli articoli disponibili online.

Molta strada è stata fatta dal primo articolo di Garfield⁸⁴ datato 1955. Mentre gli studi di qualche anno fa si riferivano semplicemente alla ricerca di una correlazione positiva tra Open Access e numero di citazioni, i lavori più recenti hanno cominciato metodicamente a dissezionare i fattori che guidano le correlazioni osservate e a scoprire cosa questo potrebbe significare per certi autori che usano il modello Green o il Gold dell'OA⁸⁵. Ciò ha comportato anche un certo fermento, proveniente dal fronte contrario all'Open Access, che ha iniziato a commissionare studi nel tentativo di smantellare le teorie dimostrate da Lawrence nel 2001 e successivamente da Harnad e Brody nel 2004.

Ma il lavoro davvero innovativo è lo studio del novembre 2006 dell'olandese Moed⁸⁶ che ha dato risultati interessanti, disponibile su *arXiv.org*. Usando il *WoS* Moed comparò - per un dato settore della fisica (condensed matter) - in un dato periodo di tempo (sette anni), tutti i lavori depositati nell'archivio dei fisici e successivamente pubblicati in un periodico, con quelli dello stesso periodico che non furono mai depositati nell'archivio. Il lavoro di Moed mette in evidenza come l'archivio acceleri il processo delle citazioni, non tanto nel numero delle citazioni prodotte, bensì come immediatezza delle citazioni, ciò in conseguenza del fatto che i lavori depositati immediatamente dopo la loro creazione accelerano il processo di conoscenza.

Moed nel suo studio fu il primo ad imporre a livello metodologico la “*variable citation window*” ovvero l'uso di finestre temporali prefissate per il conteggio delle citazioni ad ogni articolo analizzato e questo, va sottolineato, denota serietà e metodo indiscutibili.

16. Craig e i tre postulati di Kurtz

Usando la statistica citazionale, dal NASA-Smithsonian Astrophysics Data System (ADS), Kurtz e colleghi nel 2000, confermarono le scoperte degli studi precedenti. Esaminarono il tasso medio di citazione a lavori e-printed in the *Astrophysical Journal*, e dimostrarono che i lavori più importanti

⁸² The Open Citation Project - Reference Linking and Citation Analysis for Open Archives
<http://opcit.eprints.org/oacitation-biblio.html>

⁸³ Steve Lawrence, Free online availability substantially increases a paper's impact

<http://www.nature.com/nature/debates/e-access/Articles/lawrence.html>

Steve Lawrence (2001) Online or Invisible? *Nature* 411 (6837): 521

<http://www.neci.nec.com/~lawrence/papers/online-nature01/>

⁸⁴ Garfield, E. (1955) Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas
Author eprint, in *Science*, Vol:122, No:3159, p.108-111, July 15, 1955

[http://www.garfield.library.upenn.edu/papers/science_v122\(3159\)p108y1955.html](http://www.garfield.library.upenn.edu/papers/science_v122(3159)p108y1955.html)

⁸⁵ Gold sta per via d'oro e si riferisce al canale dell'Open Access che prevede la pubblicazione in periodici ad accesso aperto. Green sta per via verde, secondo canale dell'Open Access che si riferisce all'auto-deposito negli archivi aperti

⁸⁶ Henk F. Moed The effect of “Open Access” upon citation impact: An analysis of ArXiv's Condensed Matter Section
<http://arxiv.org/ftp/cs/papers/0611/0611060.pdf> (novembre 2006)

pubblicati nei principali periodici di fisica e astronomia erano stati prima sottoposti all'archivio arXiv a disposizione di tutta la comunità.

Sempre l'astrofisico Kurtz (e colleghi)⁸⁷ nel 2006, esaminava il cambiamento nel comportamento delle citazioni dal momento della comparsa sulla scena del server eprints arXiv (Ginsparg, 2001). Il lavoro degli astrofisici mirava ad avvalorare o meno la tesi che i lavori inizialmente apparsi su arXiv ottenevano maggiori citazioni rispetto a quelli che non erano stati depositati nel repository, confermando gli studi di Lawrence, 2001; Brody et al., 2004; Schwarz & Kennicutt, 2004; Kurtz et al., 2005, Metcalfe, 2005.

Nel maggio 2007 ha suscitato grande scalpore l'articolo di Craig (e altri autori) dall'eloquente titolo *Do Open Access Articles Have Greater Citation Impact? A critical review of the literature*⁸⁸, che di seguito definiremo come studio PCR. L'articolo, referato e accettato per la pubblicazione dal *Journal of Informetric*, è una rassegna critica (di parte) di alcuni studi recenti e retrospettivi focalizzati sull'eventuale correlazione tra disponibilità online ad accesso aperto di articoli scientifici e elevato tasso di citazione.

Due le questioni in gioco: la prima *"Gli articoli di ricerca ad accesso aperto hanno un impatto di citazione maggiore?"* La seconda *"I miei lavori di ricerca – e perciò io come autore – otterranno un beneficio ai fini citazionali dal canale Gold e dal Green del modello accesso aperto?"*.

Se le due domande sono legittime, proprio perché rappresentano le due fasi principali nello sviluppo della letteratura di ricerca su questo soggetto, di contro è doveroso considerare che l'articolo è stato commissionato dal consorzio commerciale *Public Research Consortium* PCR. E' opportuno evidenziare che i cinque autori provengono tutti dal settore commerciale che produce e distribuisce, a pagamento, produzioni intellettuali di ricerca e precisamente gli autori dipendono da Wiley-Blackwell, Elsevier e Thomson Scientific (WoS)⁸⁹. Non sorprende perciò che l'articolo conduca alla conclusione che la relazione tra l'accesso aperto e un alto tasso di citazione sia una leggenda da sfatare e che, alla luce di numerosi e seri lavori di analisi bibliometrica, quanto fu asserito da Lawrence nel 2001 altro non era che un abbaglio. L'articolo attacca fortemente in vari punti il metodo dell'analisi bibliometrica di Lawrence, accusandolo di ambiguità nella terminologia e di confusione nel metodo di analisi dei dati.

Peter Suber, nel suo blog⁹⁰ del 18 maggio 2007, ha commentato, in modo sottile, tali conclusioni precisando che *"we shouldn't be surprised to see that good relevant literature that is easier to find and retrieve is cited more often than good relevant literature that is harder to find and retrieve. Or, if a careful study concluded that this view is false, then one might expect it to be more careful in summarizing the reasons why."*

Ne è seguito un acceso e divertente dibattito nel Forum di discussione *American Scientist*⁹¹. Bruce Royan ha subito polemizzato sulla provenienza degli autori, ironizzando sul fatto che l'articolo sta

⁸⁷ Henneken, E. A., Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C., Thompson, D., and Murray, S. S. (2006) Effect of E-printing on Citation Rates in Astronomy and Physics
<http://arxiv.org/pdf/cs.DL/0604061>

⁸⁸ Craig, Ian; Andrew Plume, Marie McVeigh, James Pringle & Mayur Amin (2007). Do Open Access Articles Have Greater Citation Impact? A critical review of the literature. *Journal of Informetrics*.
http://www.publishingresearch.net/Citations-SummaryPaper3_000.pdf.pdf

⁸⁹ Iain D. Craig: Wiley-Blackwell; Andrew M. Plume, Mayur Amin: Elsevier; Marie E. McVeigh, James Pringle: Thomson Scientific

⁹⁰ Peter Suber (2007), Publishers doubt the OA impact advantage, *Open Access News*, 18 May 2007
http://www.earlham.edu/~peters/fos/2007_05_13_fosblogarchive.html#5358905179606249479

⁹¹ *American Scientist Open Access Forum*

<http://listserver.sigmaxi.org/sc/wa.exe?A2=ind07&L=american-scientist-open-access-forum&D=1&F=1&P=60749>

comunque “postato” ad accesso aperto sul sito di PCR “Forse probabilmente ai fini di migliorarne l’impatto?”

Ma vediamo meglio le conclusioni cui giungono i cinque autori dello studio PCR, che poggiano per certi aspetti su alcuni recenti lavori di analisi bibliometrica, in particolare lo studio Moed.

Lo studio PCR in un certo senso distorce e strumentalizza i risultati dello studio di Moed oltre a studi di altri autori, tramite un’impostazione basata sui tre postulati dell’astronomo Michael Kurtz del 2005:

- a. Il postulato Open Access (OA): suggerisce che gli autori sono maggiormente invogliati a leggere e conseguentemente citare gli articoli disponibili nel modello OA
- b. Il postulato Selection Bias (SB): suggerisce che è più probabile che la maggior parte degli autori di spicco (e perciò potenzialmente più citabili) rendano i loro articoli disponibili in un modello di OA e che mettano a disposizione i loro articoli più importanti e quindi più citabili
- c. Il postulato Early View (EV): si riferisce solo agli articoli posti ad accesso aperto prima della pubblicazione finale in un periodico (sia in versione pre-print o post-print) che rispetto all’articolo pubblicato sulla rivista ha potuto maturare un numero di citazioni maggiore, proprio a causa di questo lasso di tempo iniziale. Non riuscire a rappresentare questo effetto deve necessariamente fornire un risultato distorto.

Royan in *American Scientist* critica l’articolo PCR perché, strumentalizzando la teoria dei tre postulati di Kurtz, nel tentativo di screditare e smontare l’ipotesi che un articolo ad accesso aperto abbia maggiori possibilità di essere letto e quindi citato, di fatto ottiene l’effetto contrario.

Infatti l’articolo PCR dice che:

- Gli articoli posti ad accesso aperto entro gli archivi sono disponibili prima degli altri ad accesso chiuso che invece devono attendere i tempi di pubblicazione e quindi è per questa ragione che sono più citati.
- L’altra ragione è che gli articoli ad accesso aperto, tendenzialmente sono i migliori, perché i loro autori sono autori di prestigio.

Harnad, commentando l’articolo di PCR, concorda sul fatto che ad oggi nessuno studio è esente da difetti metodologici i quali potrebbero essere corretti, considerata la disponibilità di nuovi strumenti tecnici e le abilità metodologiche offerte dall’evoluzione della bibliometria negli ultimi anni.

Un’analisi condotta nel 2006 sulla banca dati *MathSciNet* dai bibliotecari della Cornell⁹² su 2.765 articoli pubblicati in quattro periodici matematici dal 1997-2005 indicò che gli articoli depositati nel repository arXiv ricevevano il 35% di citazioni in più di media rispetto agli articoli non depositati (un vantaggio di circa 1.1 citazione per articolo). La spiegazione più plausibile fu non tanto entro il postulato OA o EV Early View ma nel processo di Self-Selection, il quale aveva condotto al deposito di articoli di alta qualità in arXiv.

Ancora, nonostante il loro vantaggio citazionale per il fatto di essere stati resi disponibili fin da subito, gli articoli presenti in arXiv, durante i due anni successivi la loro pubblicazione su rivista, registravano il 23% di scarichi in meno (circa 10 download in meno per articolo) dal sito dell’editore, proprio perché i fisici si dirigevano in prevalenza nell’archivio aperto piuttosto che sul sito dell’editore. Sostanzialmente lo studio conferma lo stesso vantaggio di citazione riportato da altri autori, per i lavori depositati negli archivi o i qualche modo ad accesso aperto, ma questo non è tanto attribuibile all’Open

⁹² Philip M. Davis, Michael J. Fromerth, Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles? March 14, 2006
<http://arxiv.org/abs/cs.DL/0603056>

Access come causa, bensì all'auto-selezione. L'Open Access perciò potrebbe essere un risultato, non una causa, di un autore che promuove l'alta qualità.

17.Osservazioni conclusive

Nelle mie numerose peregrinazioni alla ricerca di cause, effetti, opinioni, vecchi modelli e nuove frontiere della bibliometria non ho però trovato lavori che indagassero su alcuni punti chiave che forse non sono di secondaria importanza.

Il primo punto chiave è che il mondo Open Access è aperto a citazioni verso il mondo chiuso commerciale e questo lo si vede bene nelle bibliografie degli articoli pubblicati sui periodici ad accesso aperto, ricche di link verso risorse sia del mondo OA sia del mondo editoriale a pagamento. Questo non è altrettanto vero per i periodici a pagamento basati sul modello tradizionale subscription che, nelle bibliografie a fondo di ciascun articolo, hanno link quasi esclusivamente entro la rete editoriale del modello DOI/CrossRef⁹³. Forse è plausibile pensare che questa sia una distorsione a monte del modello citazionale, quello che si potrebbe definire *one-way citation* o citazione a senso unico.

Il secondo è un suggerimento che va fatto a tutti i bibliotecari che dovrebbero capire l'importanza di inserire nei cataloghi e nei portali le risorse ad accesso aperto: periodici, archivi... questo perché se parliamo di bias non è corretto che si porgano agli utenti su un piatto d'argento solo le risorse a pagamento che in tal modo risultano più visibili, e quindi maggiormente citabili.

La terza, ma non ultima in tema di importanza è una considerazione che si riferisce alla potenza dei nuovi strumenti informativi. Se l'esplosione del Web e delle risorse informative con tutti i vari strumenti che ci rendono la vita meno complicata mette a rischio il predominio di un certo mercato editoriale che si è cristallizzato da decenni, lasciando spazio a nuovi ambiti e a nuovi modelli di indagine utili ad una comparazione metodologica più consapevole, d'altro canto questa stessa tecnologia che avanza ci pone altri pericoli. Pericoli che non sempre riusciamo a distinguere tra lo sfavillio delle meraviglie che via via ci accingiamo ad aprire. Mi riferisco per esempio alla nuova categoria di strumenti etichettati come "citation manager su web"⁹⁴ strumenti utilissimi, senza dubbio che confezionano eccellenti bibliografie in pochi minuti a corredo di un articolo appena sfornato. Strumenti in grado di convertire grandi quantità di dati bibliografici da uno stile citazionale all'altro, ne prevedono decine e decine adatti agli stili citazionali delle riviste più quotate (in termini di IF) o strutturati su set di metadati per stili citazionali personalizzati. Ma da dove attingono o diciamo recuperano i riferimenti bibliografici questi software portentosi? Dagli stessi database commerciali che ci vengono venduti e quindi non è poi tanto assurdo pensare che, bibliografie tarate non tanto su una seria selezione dell'autore, ma costruite con citazioni raccolte dai vari database del modello tradizionale, alla fine alzino il fattore di impatto di quelle risorse provenienti dal mondo chiuso, grazie a questi strumenti che costruiscono bibliografie perfette con un semplice click...

Una citazione attribuita ad Albert Einstein, recuperata dall'articolo di Ewing ci rammenta che "*non tutto ciò che può essere contato conta, e non tutto ciò che conta può essere contato*".

⁹³ <http://www.crossref.org/>

Il DOI® (Digital Object Identifier) è il nuovo standard per l'identificazione degli "oggetti digitali" creato dalla International DOI Foundation (IDF), un organismo cui partecipano tra gli altri l'Unione internazionale degli editori, alcune associazioni nazionali (l'americana, la coreana e l'AIE), aziende tecnologiche come Microsoft e Adobe e alcuni grandi gruppi editoriali multinazionali, in particolare nell'ambito professionale ed educativo. [dal sito AIE]

⁹⁴ Per un approfondimento sui *citation manager software* si rimanda al sito di Francesco dell'Orso *Personal Bibliography Management Software*, con analisi comparative e articoli vari, sul sito Burioni <http://www.burioni.it/forum/dellorso/bms/text/index.html>

Jean-Claude Guedon, nel suo illuminante saggio *“La lunga ombra di Oldenburg”*⁹⁵ ci mette in guardia sugli strumenti di informazione che, applicati ai grandi database della conoscenza scientifica, divengono potenti strumenti che possono influenzare le strategie di investimento e le politiche scientifiche nazionali e internazionali. Le possibilità strategiche di indicatori scientifici su argomenti caldi, come nuovi farmaci, o nuovi materiali o su certi progetti di ricerca, sono immense, proprio perché somigliano alle indagini di marketing applicate allo studio delle abitudini e dei profili dei consumatori, ma noi stiamo parlando di scienziati e se questi strumenti, in mano a monopoli privati, sono applicati alle produzioni intellettuali della ricerca, il rischio di uno sviluppo incontrollato di un mercato secondario di ricerche “metascientifiche” è davvero allarmante.

Strumenti utili per analisi bibliometriche sulle citazioni, sulla misurazione statistica dell’uso delle ricerche da parte dell’utente/autore, per il conseguente monitoraggio dei dati, dovrebbero essere nelle mani di chi la ricerca la produce proprio perché

“ponendo un insieme fondamentale di riviste in netto rilievo, l’ISI (e alcune altre prestigiose bibliografie) sono riuscite a trasformare la ricerca dell’eccellenza in una corsa per uno status di tipo elitista [elitario]. In se stessa, questa tendenza non avrebbe potuto destare grave preoccupazione, se il controllo delle pubblicazioni scientifiche fosse rimasto solidamente nelle mani delle comunità scientifiche, e se la spinta in senso elitistico fosse stata trattenuta entro limiti ragionevoli.”

Queste considerazioni dovrebbero far riflettere tutti i soggetti in causa, autori, esperti di bibliometria, editori, bibliotecari, al fine di condurre i futuri studi bibliometrici verso uno sviluppo coerente di metodi di ricerca sempre più robusti che tengano in considerazione eventuali distorsioni in tutti i sensi e i fattori che possono generare confusione.

18. Riferimenti bibliografici

Amelotti, Ercole. *Protocollo OAI-PMH negli Open Archive e applicazione CDSware per la rappresentazione dei relativi dati bibliografici*, tesi di laurea in informatica, Università degli Studi di Messina, A.A. 2003-2004 (relatore Puccio L., correlatore De Robbio A.).

Batista, Pablo D.; Campiteli, Monica G ; Kinouchi, Osame; Martinez, Alexandre S. Is it possible to compare researchers with different scientific interests?, *Scientometrics*, Vol 68, No. 1 (2006), pp. 179-189.

Buttà, Basilio, *Metadati negli Open Archive: tecniche di conversione dei formati dei record nel formato MARC21 utilizzato da CDSware*, tesi di laurea in informatica, Università degli Studi di Messina, A.A. 2003-2004 (relatore Puccio L., correlatore De Robbio A.)

Cassella, Maria. Novità recenti dal mondo delle statistiche di uso: il protocollo SUSHI e le nuove linee guida dell’ICOLC. *AIB notizie* 18 (2006), n. 9
<http://www.aib.it/aib/editoria/n18/0908.htm3>

⁹⁵ Guédon, Jean-Claude, *La lunga ombra di Oldenburg: i bibliotecari, i ricercatori, gli editori e il controllo dell’editoria scientifica*, traduzione italiana a cura di Brunella Casalini, Maria Chiara Pievatolo e Francesca Di Donato, “Bollettino telematico di filosofia”, 2005, <http://bfp.sp.unipi.it/hj/browse.php?browse=subjects&id=63>.

Centro universitario per la valutazione e il controllo CRESCO. Report sulla valutazione della ricerca scientifica universitaria. A cura di Maria Vella. Gennaio 2004

<http://www.cresco.unisi.it/wpeng/wp2eng.pdf>

Comba, Valentina (2003) La valutazione delle pubblicazioni: dalla letteratura a stampa agli Open Archives. *Bollettino AIB* 2003(1):pp. 65-76.

<http://eprints.rclis.org/archive/00000095/>

Craig, Ian D. Wiley-Blackwell; Andrew M.Plume, Mayur Amin: Elsevier; Marie E.McVeigh, James Pringle: Thomson Scientific

Craig, Ian; Plume, Andrew; McVeigh, Marie; Pringle, James; Amin, Mayur (2007). Do Open Access Articles Have Greater Citation Impact? A critical review of the literature. *Journal of Informetrics*.

http://www.publishingresearch.net/Citations-SummaryPaper3_000.pdf.pdf

Davis, Philip M. ; Fromerth, Michael J. Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles? March 14, 2006

<http://arxiv.org/abs/cs.DL/0603056>

Dellisanti, D. ICOLC 2006: SUSHI un nuovo strumento per la misurazione delle risorse elettroniche, *Bollettino CILEA* Nr. 105 Febbraio (2007)

http://bollettino.cilea.it/rst/rst.php?op=view_metadata&id=549

Egghe, L.; Rousseau R. (2006) An informetric model for the Hirsch-index, *Scientometrics*, vol. 69, no. 1, pp. 121-129.

Egghe, Leo; Rousseau, Ronald. *Introduction to Informetrics: Quantitative Methods in Library, Documentation and Information Science*. Elsevier, 1990. 978-0444884930

Ewing, John. *Measuring Journals. Notices of the AMS*, Vol. 53, No. 9, October 2006, 1049-1053

<http://www.ams.org/notices/200609/comm-ewing.pdf>

Figa'-Talamanca, Alessandro. *L'Impact Factor nella valutazione della ricerca e nello sviluppo dell'editoria scientifica*. IV Seminario Sistema Informativo Nazionale Per La Matematica *SINM 2000 : un modello di sistema informativo nazionale per aree disciplinar*. Lecce, Lunedì 2 ottobre 2000

<http://siba2.unile.it/sinm/4sinm/interventi/fig-talam.htm>

Garfield, E. (1955) Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas Author eprint, in *Science*, Vol:122, No:3159, p.108-111, July 15, 1955

[http://www.garfield.library.upenn.edu/papers/science_v122\(3159\)p108y1955.html](http://www.garfield.library.upenn.edu/papers/science_v122(3159)p108y1955.html)

Garfield, Eugene. The History and Meaning of the Journal Impact Factor. *JAMA*, January 4, 2006—Vol 295, No. 1 (Reprinted). <http://www.garfield.library.upenn.edu/papers/jamajif2006.pdf>

Guédon, Jean-Claude, *La lunga ombra di Oldenburg: i bibliotecari, i ricercatori, gli editori e il controllo dell'editoria scientifica*, traduzione italiana a cura di Brunella Casalini, Maria Chiara Pievatolo e Francesca Di Donato, “*Bollettino telematico di filosofia*”, 2005,

<http://bfp.sp.unipi.it/hj/browse.php?browse=subjects&id=63>.

Harnad, Stevan; Brody, Tim. Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals

<http://www.dlib.org/dlib/june04/harnad/06harnad.html>

Henneken, E. A., Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C., Thompson, D., and Murray, S. S. (2006) Effect of E-printing on Citation Rates in Astronomy and Physics

<http://arxiv.org/pdf/cs.DL/0604061>

Hirsch, J. E. An index to quantify an individual's scientific research output (2005-08-03)

In *Proceedings of the National Academy of Sciences* 102 16569 (2005)

<http://arxiv.org/abs/physics/0508025>

Kiduk, Yang; Meho, Lokman I "Citation Analysis: A Comparison of Google Scholar, Scopus, and Web of Science." *Information Realities: Shaping the Digital Future for All - Proceedings of the Annual Meeting of the American Society for Information Science and Technology (November 2006)*

<http://www.asis.org/Conferences/AM06/papers/185.html>

Koenig, M.E.D. (1983). Bibliometric indicators versus expert opinion in assessing research performance. *Journal of the American Society for Information Science*, 34, 136–145.

Lawrence, Steve (2001) Online or Invisible? *Nature* 411 (6837): 521

<http://www.neci.nec.com/~lawrence/papers/online-nature01/>

Lawrence, Steve. Free online availability substantially increases a paper's impact

<http://www.nature.com/nature/debates/e-access/Articles/lawrence.html>

McVeight, Marie

Open Access Journals in the Thomson ISI citation Databases: analysis of the impact factors and citation patterns. (Ottobre 2004)

<http://www.thomsonscientific.com/media/presentrep/essayspdf/openaccesscitations2.pdf>

Meho, Lokman I. "The Rise and Rise of Citation Analysis." *Physics World* 20, no. 1 (January 2007): 32-36.

<http://www.slis.indiana.edu/faculty/meho/PhysicsWorld.pdf>

Moed, Henk F. The effect of "Open Access" upon citation impact: An analysis of ArXiv's Condensed Matter Section

<http://arxiv.org/ftp/cs/papers/0611/0611060.pdf> (novembre 2006)

Noruzi, Alireza (2005) [Web impact factors for Iranian Universities](#). *Webology* 2(1).

<http://eprints.rclis.org/archive/00005596/>

Noruzi, Alireza (2006) The Web Impact Factor : a critical review. *The Electronic Library* 24(4):pp. 490-500.

<http://eprints.rclis.org/archive/00005543/>

Pezzilli R, Morselli-Labate AM, Morotti L, Steering Committee of "JOP. Journal of the Pancreas: Considerations on a 'Pancreatic' Electronic Journal at the Beginning of Its 8th Year of Publication"

JOP. J Pancreas (Online) 2007; 8(3):263-267. Editorial published by the Steering Committee of JoP
<http://www.joplink.net/prev/200705/index.html>

Rousseau, Ronald (2006) A case study : evolution of JASIS' Hirsch index.
<http://eprints.rclis.org/archive/00005430/>

Rousseau, Ronald (2006) New developments related to the Hirsch index.
<http://eprints.rclis.org/archive/00006376/>

Sidiropoulos, Antonis; Katsaros, Dimitrios; Manolopoulos, Yannis. Generalized h-index for disclosing latent facts in citation networks,
arXiv:cs.DL/0607066 v1 13 Jul 2006

Spigler, Renato "*Peer-Reviewing and Electronic Publishing*", High Energy Physics Libraries Webzine, issue 6, March 2002
<http://library.cern.ch/HEPLW/6/papers/5/>

Suber, Peter (2007), Publishers doubt the OA impact advantage, Open Access News, 18 May 2007
http://www.earlham.edu/~peters/fos/2007_05_13_fosblogarchive.html#5358905179606249479

The Open Citation Project - Reference Linking and Citation Analysis for Open Archives
<http://opcit.eprints.org/oacitation-biblio.html>

Trombetti, Guido Presidente della CRUI Conferenza dei Rettori delle Università Italiane, Relazione sullo stato delle Università Italiane 2006. Roma, 9 novembre 2006
www.crui.it/crui/newsletter/ottobre/lettera_CRUI_10.pdf

The Web of Knowledge: *A Festschrift in Honor of Eugene Garfield*. Edited by Blaise Cronin and Helen Barsky Atkins. 2000/544 pp/hardbound ISBN: 1-57387-099-4
<http://books.infotoday.com/asist/WebKnow.shtml>